A Meta-Analytic Critique of Mael et al.'s (2005) Review of Single-Sex Schooling

Margaret L. Signorella

The Pennsylvania State University

Amy Roberson Hayes

University of Texas at Austin

Yidi Li

The Pennsylvania State University

Author Notes

Margaret L. Signorella, Psychology Department, Greater Allegheny Campus; Amy Roberson Hayes, Psychology Department; Yidi Li, Psychology Department

Yidi Li is now at the University of Michigan.

An earlier version of this research was presented at the meetings of the American Educational Research Association, Vancouver, BC, April 2012 and the Society for Research in Child Development, Seattle, WA, April 2013.

Correspondence concerning this article should be addressed to Margaret L. Signorella, Penn State Greater Allegheny, 4000 University Drive, McKeesport, PA, 15132-7698. Email: msignorella@psu.edu.

The final publication is available at:
http://link.springer.com/article/10.1007/s11199-013-0288-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,* 423-441. doi: 10.1007/s11199-013-0288-x

Abstract

Continued controversy in the United States over the relation of school gender composition to student outcomes is partially due to the failure to reach consensus on how to evaluate the research and the conclusions to be drawn from the research. In this review we examine major problems in the available research on single-sex schools, in terms of both theory and methodology. We critically examine the major review done by Mael, Alonso, Gibson, Rogers, and Smith (2005) and identify serious problems in studies under review and the review methods. Drawing from both U.S. and non-U.S. samples, a subset of outcome measures (mathematical performance, verbal performance, self-esteem or self-concept) were analyzed using meta-analytic techniques. No significant effects for school gender composition were observed for either mathematical performance or self measures. There was a small advantage observed on verbal performance for students in single-sex schools, but in this comparison there was a strong relation between the direction of the effect on the outcome measure and pre-existing differences. For the three comparisons combined, preexisting student differences in SES or cognitive performance revealed advantages to students in single-sex schools. In conjunction with research showing that control of preexisting differences reduces or removes single-sex school advantages, this review provides support for the position that single-sex schools may offer no advantage to student cognitive performance or social development, but rather may reflect the consequences of a more select and advantaged student population.

Keywords: single-sex schooling, gender differences, meta-analysis

A Meta-Analytic Critique of Mael et al.'s (2005) Review of Single-Sex Schooling

Introduction

The United States has seen a dramatic increase in the number of public single-sex schools in the past decade since the passage of amendments to Title IX (Bigler & Signorella, 2011). According to the National Association for Single-Sex Public Education (NASSPE), there were 116 single-sex public schools and 390 coeducational public schools with single-sex classes in the United States in the 2011-2012 academic year (NASSPE, 2012). Creating single-sex schools, especially in the public arena, can be difficult and disruptive for school districts (e.g., Chute, 2012). It is important, therefore, that scholars carefully examine the costs and benefits of separating students by gender for education. The No Child Left Behind Act approved funding for innovative education programs, including single-sex schools and single-sex programs within coeducational schools. Districts were to conduct self-evaluations of their single-sex classes at least every two years and ensure that a "substantial relationship" exists between the single-sex nature of the classes and achievement of the schools' educational objectives (U.S. Department of Education, 2006). Reviews of the empirical literature to date, however, have produced no apparent consensus on outcomes (e.g., Bracey, 2006; Haag, 1998; Marsh, 1989; Mael, Alonso, Gibson, Rogers, & Smith, 2005).

The United States Department of Education commissioned a review of the single-sex education evidence from the U.S. and abroad as part of the process of changing the regulations, and subsequently published a report based on those findings (Mael, Alonso, Gibson, Rogers, & Smith, 2005). Using a vote-counting method to code significant findings from empirical studies, Mael et al. (2005) concluded that the evidence for single-sex schooling remained equivocal, but also seemed to argue that there was evidence of a single-sex advantage (e.g., p. 83, p. 84, p. 85, p. 86). Since its release, Mael's report has been cited frequently in the single-sex education debate (e.g., Chadwell, 2010; Smithers & Robinson, 2006; Kessels & Hanover, 2008). However, the methodology employed by the authors of the 2005 report, as well as the methodology of the studies analyzed therein, represent a myriad of methodological and interpretational challenges. The purpose of the present paper is to provide an overview of the methodological and data analytic issues in this disputed literature, and to illustrate some of the problems by reexamining selected findings from Mael et al.'s (2005) report using meta-analytic, rather than vote counting, methods. First, we examine the methodological and theoretical shortcomings present in the single-sex education literature as well as in Mael et al.'s (2005) review of the literature. Next, we re-analyze the larger groups of studies used in Mael et al.'s (2005) report using meta-analysis. Finally, we will discuss the implications of our findings for the single-sex education debate in the United States.

Problems in the Single-Sex Education Literature

Several limitations characterize the single-sex education literature: (a) lack of theoretical grounding, (b) lack of sufficient controls in methodology and study design, and (c) weakness of the strategies used to summarize existing data.

Theoretical issues. The first major shortcoming of the literature on single-sex schooling is the lack of theoretical grounding, which then impacts the methodological decisions made. Proponents of single-sex schools purport a diverse range of benefits of this educational strategy, which are hypothesized to include academic, psychological, and behavioral advantages (Sax, 2005; Chadwell, 2010). As briefly summarized by Bracey

(2006), single-sex education is perceived as a means to increase the representation of girls and women in traditionally-masculine fields (science, technology, engineering, math), to control boys' behavior problems, to remove the heterosexual distractions of coeducational schools (thus increasing the academic motivation of both genders), and to improve girls' self-esteem. Mael (1998) discussed the supposed merits of mixed-sex (also called coeducational or CE) environments by proponents and the resulting listing was short (e.g., more like real-life, less stereotyped, and cheaper) and included a nonbenefit: "CE advocates do not typically claim that coeducation is pedagogically superior" (p. 103). The many more proposed benefits for single-sex education listed by Mael (1998) included a mix of claims: that single-sex education is more natural (e.g., everyone prefers it, most societies separate by gender for education, boys and girls are different on a variety of dimensions), and that it reduces stereotyping in girls, reduces the feminization of boys, and improves the academic atmosphere (pp. 104-105).

Many of these hypotheses about the effects of single-sex schools, however, are proffered with insufficient, or nonexistent, theoretical grounding (see Halpern et al., 2011). The majority of the research on single-sex education focuses on the proposed outcomes of single-sex education cited above without attention to the mechanisms underlying the purported effects. Furthermore, it is crucial to distinguish between hypothesized effects that are centered on theoretical mechanisms and those that may be accounted for by third variables that are correlated with both the gender composition of the school and the outcome measure. Many have argued that incoming differences between students in single-sex schools and mixed-sex schools, such as socioeconomic status or parent involvement, may be responsible for the positive effects of single-sex schools observed in some studies

(Marsh, 1989; Riordan, 1994). A few researchers, including Riordan (1990; 2002) have argued that there may be a different set of expectations in single-sex schools that make them more effective in comparison to traditional mixed-sex schools. Although these comparisons are instructive and will be discussed in more detail below, they have little to do with the impact of school gender composition on student outcomes. For example, if academic performance is hypothesized to be higher in single-gender environments, a potential mediator might be that students are less distracted in single-gender classrooms. Only a few researchers have tested reasonable, theoretically-grounded hypotheses for why educating boys and girls separately would make a difference in student outcomes. However, a majority of studies comparing the two types of schools are primarily concerned with outcomes like academic achievement without any measure of school factors (e.g., teacher quality, student selection, financial resources), other than gender composition, that contribute to the outcome.

Finally, some of the explanations for the benefits of single-sex education are contradictory. Does single-sex education reduce stereotyping for girls or does mixed-sex education? Is it considered valuable to reduce stereotyping for girls but increase it for boys? Does all academic performance increase in single-sex schools, as argued by those who say that single-sex environments are more academically oriented, or does it reduce gender differences in specific areas, such as math and language? If all students perform better in every subject in a single-sex environment, then specific gender differences would presumably remain, unless one argues that that the overall increases in performance will be accompanied by greater increases in gender-stereotypical domains by the group that is presumably behind.

Methodological issues. The second major shortcoming of the literature concerns the methodologies typically employed in the field. The "gold standard" for assessing causal effects of school programs is random assignment and blind assessment, and no studies employing these techniques exist. Federal regulations require that enrollment in single-sex settings be voluntary (U.S. Department of Education, 2006), and thus truly randomized designs are impossible to implement in the U.S. Nonetheless, the weakness of non-random assignment might be addressed via methodological procedures and/or statistical controls. The use of such designs and controls are crucial because much of the reported success of single-sex education may be attributable to selection and school quality effects.

Selection effects. Researchers cannot take for granted the similarity of single-sex schools due solely to their similar gender compositions, just as one would not assume that all coeducational schools are "the same" because they have a mixture of boys and girls. Within the population of single-sex schools, schools differ both in terms of the students (and/or their families) who elect to attend, and who are selected by the schools to attend (Hayes, Pahlke, & Bigler, 2011). First, students who elect to attend single-sex schools may differ systematically from those students who do not attend (Riordan, 2002). The problem of preexisting differences between students who attend single-sex schools versus those in mixed-sex schools was a key factor in the disagreements over interpretation the High School and Beyond survey data (e.g., Lee & Bryk, 1986; Marsh, 1989), a massive US cohort sequential study starting in 1980 with 10^{th} and 12^{th} graders and also tracking the students after high school graduation (e.g., Lee & Marks, 1990; for an overview, see National Center for Education Statistics, n.d.). Marsh argued that the lack of information about the students in the High School and Beyond sample prior to the start of the data collection was a major problem in interpreting school effects (e.g., Marsh, 1989), and also noted that an anticipated newer longitudinal data collection in the U.S. (called NELS:88) might help answer some of these questions because the new data collection included student performance prior to high school (Marsh, 1991, p. 349; for an overview see National Center for Education Statistics, n.d). Subsequently, LePore and Warren (1997) analyzed these new data and showed that there were no significant differences in change over time that could be attributed to school gender context (p. 505), once other student background factors were taken into account statistically.

Second, those applicants who are selected by the administrators to attend single-sex schools may differ systematically from those who are not selected. Hayes et al. (2011) argued that all single-sex schools have selection criteria, such as admissions standards, and that even public single-sex schools employ selection, as most are charter or magnet-type schools that require applications. In their recent study, Hayes et al. (2011) showed that there were incoming differences in academic achievement between students who were selected for a public single-sex school compared to those who were rejected.

Cultural context of schools. Additionally, the national and cultural context of the single-sex and mixed sex schools must be acknowledged when making any comparisons among the results of different studies. Much of the research on single-sex education has been done outside the U.S., including many of the samples analyzed in the current paper. Public single-sex education has a long history in several countries, including Japan, Thailand, Ireland, the United Kingdom, Belgium, Nigeria, Australia, and New Zealand (Carpenter & Hayden, 1987; Robinson & Smithers, 1999; Mael et al., 2005). Often, these studies of single-sex education outside the U.S. have been used as support for implementing similar programs in the United States, even though the education systems in these comparison countries differ in significant ways from the U.S. system. As we will discuss below, there are also many single-sex schools world-wide that are private, or non-government schools. Importantly, there are clearly documented differences between government and private or religious-based schools that can be confounded with the gender composition of the school. We include a discussion of these differences below in order to highlight the heterogeneity of the school contexts included in the Mael et al. (2005) review.

In Australia, for example, the majority of single-sex schools are in the private, or non-government sector of schools (Australian Bureau of Statistics, 1997). Furthermore, students attending schools in the private sector in Australia generally come from more privileged socioeconomic backgrounds than students attending the government schools (Australia Bureau of Statistics, 1997). The 1997 report also identifies another difference; namely, that government schools are larger than non-government schools, which could also affect student outcomes. There are also differences between states within Australia. Carpenter and Hayden (1987) contrast Victoria, where "there is a tradition of single-sex, public secondary schools, and though their numbers have diminished, many of these schools still exist" (p. 158), with Queensland, where "all public secondary schools are coeducational" (p. 158). Thus, it would be difficult and complicated to draw any type of comparison between Australian single-sex schools as a group and public or private single-sex schools in the United States.

England has seen a large decrease in the number of single-sex schools since the 1960s (Robinson & Smithers, 1999). Similar to Australia, there is "a concentration of single-sex schools in the private and selective sectors" (Sullivan, 2009, p. 265). Smithers and

Robinson (2006) concluded that there are differences among types of schools in England beyond gender composition that could account for observed outcomes, such as the apparent advantage to single-sex schooling for girls, but not boys, reported in Spielhofer et al. (2002, 2004). A related problem originally identified by Speilhofer et al., and expanded upon by Smithers & Robinson (2006), is that:

> girls' schools [are] over-subscribed with their pupils sometimes travelling a considerable distance to attend suggesting 'that they are deliberately chosen by parents who are informed and interested in their children's education.' Or to put it another way, they [meaning the schools] will have to select their pupils, if not academically then on some other criterion (p. 7).

A cross-cultural study by Baker, Riordan, and Schaub (1995) compared countries that varied in terms of the frequency of single-sex schools. Baker et al. (1995) used Japan as an example of a country that has relatively few single-sex schools. In Japan, elementary and junior high education is compulsory, while high school education is not (Cooke, 2005; Frasz & Kato, 1999). Additionally, high schools in Japan are more likely to be private than the compulsory lower schools (Cooke, 2005). Another country used in the Baker et al. (1995) study was Belgium, which at that time had larger numbers of single-sex schools, and has seemingly maintained a tradition subsequently in the private sector in which "some schools continue to stress their distinctive character as a girls' or boys' school, thereby discouraging prospective pupils of the opposite sex from applying" (Brutsaert, 2002, p. 73).

The above examples indicate that there are many differences across countries in school systems, including in the frequency with which they offer single-sex schools; whether those single-sex schools are public, private, or both; and how the schools are perceived within each country. It appears that within many countries there is reason to believe that there are preexisting differences between single-sex and mixed-sex schools that would predict greater success among students in single-sex schools. Without controlling for student background characteristics, therefore, international researchers may succumb to the same statistical pitfalls of U.S. researchers in comparing differentially selective schools. Accordingly, any use of international data in the U.S. single-sex schooling debate also must be subjected to rigorous methodological scrutiny.

Weak review methods. The third problem with the existing reviews of the single-sex educational literature is that the reviews are largely narrative (see Salomone, 2003, Haag, 1998; Bracey, 2006; Smithers & Robinson, 2006; Thompson & Ungerleider, 2004). One of the reasons meta-analysis has gained support as a literature review technique is that narrative reviews can be interpreted differently by different readers and may rely on misleading vote-counting methods. Borenstein et al. (2009) openly labeled vote counting as "wrong" (p. 252) and a method that is "more misleading as the amount of evidence (the number of studies) increases" (p. 251). Bushman and Wang (2009) agree that traditional vote-counting can be fairly described as "crude, flawed, and worthless" (p. 208). Other problems with vote-counting review methods are that they do not consider sample size, do not allow for determination of effect size, and offer little power to detect small to medium group differences (Bushman & Wang, 2009, p. 209). Borenstein et al. (2009) further argued that presenting either vote-counting or narrative summary reviews can lead to the same problems:

> Many statisticians would argue that, when faced with a series of studies, people have an almost irresistible tendency to draw some summary conclusions from them.

> Experience has shown that seemingly intuitive ad hoc summaries...are often highly misleading. This suggests that a statistical summary with known, but perhaps poor, properties (such as high uncertainty) may be superior to inviting an ad hoc summary with unknown properties (pp. 363-364).

To date, no one has done a meta-analysis of the single-sex education literature. Mael and his colleagues (2005) concluded that a meta-analysis of single-sex school data was not possible due to Department of Education rules on using only experimental studies, and instead used the traditional type of vote-counting that has major problems. Bracey (2006) also counted votes in comparison to some of the Mael et al. (2005) conclusions in addition to his narrative review. Beyond the statistical issues identified above, there is an additional serious problem with the vote counts in Mael et al. (2005); specifically, many of the votes in favor of single-sex education benefits come from the same authors or from authors using the same data set (Smithers & Robinson, 2006).

The Present Review

In the critique and meta-analysis to follow, we review the set of studies featured in Mael et al.'s (2005) report, and use the comparisons that have sufficient numbers of effect sizes to illustrate the difference in conclusions between narrative and quantitative reviews. In these comparisons, we will also examine the relationship between the size and direction of effects in student outcome measures with the size and direction of preexisting differences between students in single-sex versus mixed-schools.

Method

Sample of Studies

The sample used in this meta-analysis was identified by Mael et al. (2005) and was originally comprised of 40 articles, papers, and dissertations. Specifically, these studies were referred to in Mael et al. (2005) as "Quantitative Phase III-Coded references" (pp. 94-97), and we have listed them in Appendices A and B (available in online supplementary materials). Appendix A shows most of the studies in order as they are listed in Mael et al. (2005, pp. 94-97) with the following exceptions. First, articles with various dependencies are grouped together. Second, the seven studies using the High School and Beyond database are listed separately in Appendix B. Mael et al. (2005) also reviewed a small number of qualitative studies, also described below.

The original search by Mael et al. (2005, pp. 3-4) followed usual procedures for a review or meta-analysis, in that databases were searched, review articles were scanned, and then citations were followed back via the *Web of Science* citation index (then called the *Social Science Citation Index*).

Mael et al. (2005) Selection Criteria

Because this is a meta-analysis on research already selected by other authors, we therefore will review the selection criteria in the Mael et al. (2005) review. Also, because we have several disagreements with the way the studies were classified, we summarize the original process in some detail. The original selection criteria used by Mael et al. (2005, pp. 4-5) were studies that had full-time students from elementary, middle or high schools (i.e., no summer camps or special programs), were from "Westernized countries" (p. 4), were reported in English, and used single-sex comparison schools that were completely segregated by gender. Of the 2221 abstracts retrieved by Mael et al. (2005) from the initial search, 379 appeared to meet the selection criteria. The numbers of abstracts were then reduced to 102, based on the presence of control for preexisting differences (Phase II).

Among the 102 remaining studies, 88 were characterized as quantitative and 14 as qualitative (p. 5).

Next, the complete text of the 102 Phase II studies were reviewed, and Mael et al. (2005) excluded studies that had, in their view, a variety of problems, such as problems in not adequately meeting the earlier specified selection criteria, such as age of students, or problems in the methods or analyses. Each excluded Phase II study and the exclusion reason were listed in Mael et al.'s (2005) Appendix 3 (pp. 113-115). We reviewed this list of excluded studies and found that there are several whose exclusion seemed inconsistent with the stated criteria. For example, studies by Gilroy (1991), Mensinger (2003), Schneider, Coutts, and Starr (1988) and others appear to have the needed controls. We also note that other studies often cited in the single-sex literature are completely missing from Mael et al.'s review (e.g., Finn, 1980; Granleese & Joseph, 1993), including the original study (Dyer & Tiggemann, 1996) on which an included study (Mensinger, 2001) was based.

Mael et al. also excluded 10 of the 14 qualitative studies (see pp. 101-102), but did not similarly specify the reasons for each article's exclusion. Instead, the authors gave general reasons for the exclusion decisions. One, "methods used were largely quantitative" (p. 9), was confusing because the authors did not explain why the study or studies were not moved back into the quantitative group and thus included in the review. Another, "they focused on gender equity and gender consciousness issues, not on the efficacy of single-sex schooling instruction" (p. 9), was puzzling because many of the hypotheses about the impact of single-sex schools on instructional outcomes clearly suggests that gender equity is a moderator of those effects. We wish to disclose here that one of us (Signorella) co-authored an article (Signorella, Frieze, & Hershey, 1996) we believe is clearly quantitative, but that is on the excluded qualitative list (Mael et al., 2005, p. 102). As only four qualitative studies remained after the exclusions, we will not discuss these further.

Coding of the remaining 40 Quantitative Phase III articles by Mael et al. (2005, p. 6) followed standard procedures with multiple coders and discrepancies resolved by discussion. There were no data reported on the degree of agreement among coders, other than occasional references to unusual coding decisions made on individual studies (e.g., Mael et al., 2005, p. 17, p. 23).

Meta-analytic Procedures in the Present Review

For our meta-analyses, we first reviewed 39 of the 40 Mael et al. (2005) Quantitative Phase III-Coded references. The one exception was an AERA paper presentation that we could not find (Daly, Ainley, and Robinson, 1996, cited in Mael et al., 2005, p. 95). The items we coded in the remaining 39 papers were the sample characteristics (gender, age, ethnicity, country), school characteristics (public, private, religious), dependent and control variables, and results. These details are presented in the tables that follow. Four of the articles were used for training in the coding system, and the rest were coded independently. The first author coded all articles and almost all (82%) were coded by one or both of the other authors. Interrater reliabilities (kappas or intraclass correlations) ranged from 0.79 to 1.0, with a median of 0.91. Most of the disagreements occurred on the description of the school type (i.e., private, public, etc.). Disagreements were resolved by discussion.

The effect size used was the standardized mean difference (*d*), corrected for bias in the analyses (Hedges's *g*) (Borenstein et al., 2009, pp. 26-28). We have coded the effect sizes so that positive *g*s represent an advantage to single-sex schools and negative *g*s an advantage to mixed-sex schools. The standardized mean difference is computed from *M*s, *SD*s, and *n*s; many studies, however, did not report that information for the desired comparison. There are several other statistics that produce an algebraically equivalent *d* to that computed from *M*s and *SD*s, such as *r*, *t*, and *F*s from some ANOVAs (e.g., see Lipsey & Wilson, 2001; Shadish, Robinson, & Lu, 1999). Good approximations can be obtained from transformed 2 X 2 frequency data and a *p*-value with sample size (Lipsey & Wilson, 2001; Shadish et al., 1999). A verbal statement that a difference is or is not significant is hugely problematic for several reasons, and thus we will only list this information in our tables.

Comprehensive Meta-analysis (CMA) was used for both effect size computation and any further meta-analytic procedures, with additional cross-checks on effect size computations with two other effect size computation programs (Shadish et al., 1999; Wilson, 2010). The random effects model was used in our analyses as it seemed the most appropriate for the data set, given the wide variety of measures, types of schools, and other study factors included in each comparison (Borenstein et al., 2009, pp. 83-84). In addition, the characterization of this set of studies as either complete or the best is clearly not supportable, based on the problems and omissions we detail. Finally, the significance of the combined effect sizes and their homogeneity, as well as the examination of moderators, were all computed with CMA.

An additional complication in these data is that the studies varied in how they applied controls to the outcomes. Some controls were implicit, as a result of the authors asserting comparable SES and related characteristics across schools (e.g., Bornholt & Möller, 2003). In other studies, researchers applied controls through a variety of statistical means (e.g., multiple regression, ANCOVA, HLM, gain scores) and the controlled outcomes were reported in a variety of ways. Recently, Peterson and Brown (2005) have argued that βs from multiple regression between -.50 and .50 can be used to estimate *r* and used in a meta-analysis. The recommended conversion is $r = \beta + .05\lambda$, where $\lambda = 0$ if β is negative and 1 otherwise (p. 179). We therefore used βs for this purpose, when βs were used to represent either uncontrolled or controlled comparisons.

Although we report both uncontrolled and controlled comparisons in the tables for each analysis, there were many fewer available controlled comparisons where an effect size could be calculated. Most problematic were those studies in which gain scores were used to measure the effect of school gender composition. Gain scores cannot appropriately be converted to effect sizes without the intraclass correlation data (e.g., Shadish et al., 1999), and that information is virtually never reported. This means that our overall analyses will be limited to the uncontrolled effect sizes and will have implications to be discussed later.

Results

Preliminary Assessment of Quantitative Articles in the Present Review

Out of the Mael et al. (2005) original list of 40 studies (see Appendices A and B), we were, as noted above, able to retrieve 39. Among the remaining 39 studies was a Daly and Ainley paper labeled as "n.d." with no other citation information beyond the title. We found a published article with the same authors and the same title, and because the Mael et al. (2005) description of the study matched, we felt safe in assuming the published article (Daly & Ainley, 1999) is the same one reviewed by Mael et al. (2005). We searched for any additional information on the 39 studies, and as shown in Appendix A (in online supplementary materials), we found that Sanders (1992) published his report in 1995

(Sanders & Reed), and Spielhofer et al. (2002) was published in 2004. Mensinger (2001) was a re-analysis of a study by Dyer and Tiggemann (1996). Unfortunately, this search for additional information was not particularly helpful.

There are several issues to note concerning these studies featured in Mael et al. (2005), supposedly the best of all research on single-sex schools as of the mid-2000s. Together, the High School and Beyond (HS&B) database studies and 34 samples from other sources make a total of 35 independent samples. Of the 35 independent samples, 24 (69%) of the data sets are from outside the U.S. In addition, almost all of the public school single-sex samples are from outside the U.S., as all of these studies were done prior to the revision in the US Department of Education regulations. Although it may be somewhat difficult to see in scanning the appendices, most of the studies focus on academic performance or on self-esteem/self-concept. It should be easy to see, however, that many of the studies assess multiple dependent variables, and thus end up being represented in multiple comparisons in the Mael et al. (2005) review. Finally, many of these samples (e.g., the HS&B sample) have very large *N*s, and thus there may be statistically significant differences of very small magnitude.

Non-independence of samples. One of the serious problems in this set of studies is the issue introduced above; namely, that many of the 40 reports from Mael et al.'s (2005) original list were not independent. One of the assumptions of any meta-analysis is that the effects being compared are from a different set of participants. The common dependencies encountered in meta-analysis occur when the same participants are given multiple measures, a problem here as shown clearly in the appendices. A more subtle form of dependency is when studies from the same research lab or author are included in the sample. Much attention is devoted in the meta-analysis literature to solutions to these dependency issues (Borenstein et al., 2009), and the most common are averaging across measures if possible, or making a selection either randomly or based on some stated criteria (e.g., Lipsey & Wilson, 2001). The former method is usually problematic because to do properly requires that there be information about the correlation between the measures being averaged (Borenstein et al., 2009), and none of the studies in this collection have that additional information available. As a result, we will be handling dependencies by selecting a measure for inclusion based on stated criteria to be specified below.

There are two sets of studies with large to complete overlap (Caspi, 1995 and Caspi et al., 1993; Harker, 2000 and Harker & Nash, 1997). In the former instance, Caspi (1995) wrote a book chapter based on the Caspi et al. (1993) article. The book chapter does have one additional set of measures not reported in the article. In an earlier narrative review of this literature, Mael (1998) recognized that Caspi (1995) and Caspi et al. (1993) were based on the same participant sample (Mael, 1998, p. 115). In the latter instance, Harker and Nash (1997) is a conference presentation, and Harker (2000) is the apparent resulting publication. As the data tables in the two papers are identical, it seems a safe assumption that these two are the same report on the same sample.

Another form of the dependency problem is somewhat less problematic. Daly's studies (1995; 1996; Daly & Shuttleworth, 1997) appear to be based on different comparisons from the same two data sets (Survey I & Survey II), with the Daly and Shuttleworth report adding a third sample (Survey III) to the comparisons. It does not seem that there is much overlap in the specific comparisons being reported across the three articles, but it is still important to acknowledge the apparent dependencies among the data sets.

A third type of dependency present in this sample of studies is that many authors have used the same publicly available dataset to address the issue of single- versus mixed-sex schools. The articles by Lee and Bryk (1986, p. 383), Lee and Marks (1990, p. 580), Marsh (1989, p. 74; 1991, p. 332), Riordan (1990, ch. 5; 1994, p. 186), and Thompson (2003, p. 263) all use the same High School and Beyond (HS&B) database, and are listed separately in Appendix B (see online supplementary materials). We find this mistake particularly difficult to understand. The debate between Marsh (1989, 1991) and Lee & Bryk (1986, 1989) over the interpretation of the HS&B dataset is a classic in the single-sex schooling literature. The authors of the articles and chapters themselves clearly identify the source of their data and reference one another as they discuss the differing interpretations of the data. Earlier reviewers have correctly identified these various re-analyses of the HS&B data, including a 1993 Department of Education review by Moore, Piper, and Schaefer (1993) and the narrative review by Mael (1998, e.g., pp. 107-108). Although Smithers and Robinson (2006) also evidenced trouble keeping the various datasets straight, misidentifying Riordan (1985) and LePore and Warren (1997) as using HS&B, their general point that there is "multiple counting" (p. 7) of the same data is still correct. Furthermore, HS&B was an enormous study with many, many dependent measures, so the influence of this one dataset on the Mael et al. (2005) tabulation is potentially hugely problematic.

The other, less serious use of a duplicate dataset is the National Longitudinal Study (NLS) of 1972, used by Riordan in 1985 (p. 524) and 1990 (Ch. 6), and carefully and clearly identified as such by Riordan in both instances. As noted above, Smithers and Robinson

(2006) misidentified Riordan (1985) as from HS&B. The outcome data from NLS on which Riordan is reporting in 1990 are longer-term impacts, so the dependent measures are different ones than he examined in 1985.

Errors in the tabulation of results. We believe there are errors in the tabulation of results in Mael et al. (2005). We will report the exact pages of each article from which we identified what we think are errors in Mael et al.'s (2005) summary of the article so that others can evaluate our conclusions. First, there are several summaries of results that are inaccurate in various ways, as other authors have also identified (Bracey, 2006; Smithers & Robinson, 2006). For example, results from Carpenter and Hayden (1987) are mischaracterized as representing a SS versus CE comparison (Mael et al. 2005, p. 13), when they are actually a comparison of combined SS and CE results from two states in Australia: Queensland and Victoria (Carpenter & Hayden, 1987, p. 159 for sample description; p. 161 Table 2 for data).

The verbal summaries (labeled "Study Findings" in Mael et al., 2005, Tables 2-33) and the overall conclusions (labeled "Advantage to:" in Mael et al., 2005, Tables 2-33) do not always match. Mixed results (i.e., those with no consistent direction of effect) are not always classified the same way across comparison tables (also noted by Bracey, 2006). Dependent measures are omitted from summaries for no apparent reason (see Appendices A and B for listings of these omissions), or placed in the wrong category (also identified in the Appendices). As individual results are presented, these issues will be discussed further, and identified where relevant in table notes.

Example analysis illustrative of issues in Mael et al. (2005). To illustrate further some of these issues, we will use the set of studies labeled by Mael et al. (2005) as "all-subject achievement test scores" (p. 13), defined as "scores indicating the composite mastery of specific skills" (p. 13). Mael et al.'s Table 2 (2005, pp. 13-17) lists nine studies in this category, one of the relatively larger numbers of samples for a particular type of dependent measure. The first error in the Mael et al. (2005) compilation is that the Carpenter and Hayden (1987) study, as described earlier, is not correctly summarized. Carpenter and Hayden reported comparisons from two states in Australia, each with both single-sex and mixed-sex school types. The summary reported in Mael et al. (2005, p. 13) is actually the comparison of scores between the two states, collapsed across school gender composition (see Carpenter & Hayden, 1987, Table 2, p. 161), but incorrectly represented by Mael et al. (2005) as a school gender composition comparison.

We also disagree with the classification of the dependent measure in one of the two states examined in Carpenter and Hayden (1987). Only one of the two states (Victoria) reports data on all-subject achievement test scores. The other state, Queensland, reported grades rather than test scores (Carpenter & Hayden, 1987, Table 1, p. 160), a dependent measure included by Mael et al. in a different comparison (2005, p. xiii, p. 32, p. 77).

As it turns out, the overall conclusion in Mael et al. (2005, Table 2, p. 13) could be viewed as correct (advantage to SS females), because the one dependent measure that is an overall achievement score (from the state of Victoria) does show better performance by girls in single-sex schools. But what puzzles us is that if one accepts the apparent judgment by Mael et al. (2005) that the Queensland results for grades are a comparable measure to the achievement score from Victoria, then there is no consistent advantage to single-sex schools because there is no effect for school gender composition in Queensland. Specifically, there is a tiny, nonsignificant zero-order correlation (r = .032, n = 460), and no significant effect in the regression (β = .031), as shown in Tables 2 and 4, respectively, in Carpenter and Hayden (1987). The text summary is quite confusing and explicitly cites the Queensland regression as showing that "sex composition of schools did not predict (no) all-subject achievement test scores" (Mael et al., 2005, Table 2, p. 13). Thus, it seems that a more accurate overall summary (in the "Advantage to:" section) would be mixed rather than single-sex advantage.

The second problem in the all-subject achievement comparison is that there are three High School and Beyond studies out of the nine total listed (Lee & Bryk, 1986; Marsh, 1989; Riordan, 1994). Only the study by Lee and Bryk (1986) is explicitly identified as coming from HS&B (Mael et al., 2005, p. 15). Additionally, we are puzzled by the inclusion of the Lee and Bryk (1986) and Marsh (1989) studies in this section because neither report "all-subject achievement test scores." The text summary of Lee and Bryk (Mael et al., 2005, p. 15) does suggest that an "all-subject" comparison was drawn from a summary of all achievement scores reported, which would indeed overlap with the later specific comparisons (e.g., math, verbal). For the Marsh (1989) summary (Mael et al., 2005, p. 15) it is not clear what measure or measures are being reported. Riordan (1994, p. 195) does use a composite score, and the Riordan summary (Mael et al., 2005, p. 15) accurately reflects the findings, although this still presents overlap with the other HS&B samples and the individual math and verbal results. The most puzzling aspect of the HS&B study presentation is that there is a footnote to Mael et al.'s Table 2 (2005, p. 17) in which Marsh (1989) is chastised for using too many controls in his study, but at the same time fails to identify Marsh (1989) as performing an explicit reanalysis (using those additional controls) of Lee and Bryk (1986).

Finally, there are two other studies for which effect sizes could not be computed (Caspi, 1995; Spielhofer et al., 2002, 2004). Both are reporting on comprehensive exams of a type not typically used in the United States. Caspi (1995) is a second report on the same sample of girls from Caspi et al. (1993), in which the additional data on exams scores are reported. Similar to the results for social factors also examined by Caspi et al. (1993; Caspi, 1995), there were complex relations among age at menarche and school gender composition, but not enough information to compute the unadjusted effect sizes. The problems with the Spielhofer et al. (2002, 2004) study have been discussed above; namely, there was no control for factors that could have resulted in the girls in single-sex schools being a particularly selective group, and thus accounting for the better performance by girls (but not boys) in single-sex schools. It is also a problem in this latter study that, in the absence of any effect size information, a significant difference in a sample in the hundreds of thousands is difficult to interpret.

Specific Meta-Analytic Comparisons

Based on the availability of a critical mass of studies, we were able to use some of the original Mael et al. (2005) comparisons in a meta-analysis. As can be seen in Mael et al. Table 1 (2005, pp. xiii-xiv; reproduced as Table 34, pp. 77-78), there are many types of outcomes listed under some general categories, but there are few outcome types for which there are more than a couple of available studies. As will be described below, there are three types of outcomes in which there are sufficient numbers of independent effects to make a reasonable meta-analytic comparison: mathematics achievement, verbal achievement, and self-concept/self-esteem.

The first general category discussed by Mael et al. (2005), "Concurrent, Quantifiable

Indicators of Academic Accomplishment," (p. 12) has three outcomes with an apparently reasonable number of studies besides the all-subject achievement group that we showed could not be used; specifically, math, science, and verbal. There are, however, fewer science effects and many are from the same datasets as the math results. Thus, we will focus on math and verbal with the largest *n*s of independent effects in this first group of studies.

The only other general section with reasonable numbers of studies is the third one, called "Concurrent Quantifiable Indicators of Individual Student Adaptation and Socioemotional Development" (Mael et al., 2005, p. 36). There are 14 articles listed in a group called "School Track and Subject Preference" (Mael et al., 2005, p. 77). Similar to other comparisons, there are duplicate datasets reported separately, artificially inflating the total. A new problem in this particular subset is that some of the studies listed in this section do not seem to fit with the others. Several studies look at actual course taking patterns (e.g., Ainley & Daly, 2002). It is in this section that the one missing study (Daly, Ainley, & Robinson, 1996, cited in Mael et al., 2005) is located, but even if we could include this study there would not be enough independent effect sizes to combine. The other studies assess students' attitudes toward school in general or toward specific school subjects (e.g., Sanders, 1992). These all do not seem to belong together, and there are not enough in any one group to meta-analyze. The next highest numbers of available studies are self-concept at seven and self-esteem at six (Mael et al., 2005, p. 77). Self-concept and self-esteem are close enough concepts that we feel we can combine them into one analysis.

Mathematics achievement. For this analysis, the relevant studies and data are shown in Table 1, along with the effect sizes for school gender composition on mathematics achievement before and after applying controls, if possible. Also shown are the effect sizes for school gender composition on the two most common control variables reported, SES (variously measured) and prior ability (variously measured). All computable effect sizes for either dependent measure or control variables are expressed as Hedges's *g*, with positive values indicating an advantage to single-sex schools, and negative values indicating an advantage to mixed-sex schools. We will focus on the effect sizes reported prior to the application of any controls, as there are more available, they are the typical effect size used in a meta-analysis, and they also provide the most conservative test of the impact of single-sex schools.

Mael et al. (2005) originally identified 14 studies in the mathematics category, some of which proved to have independence issues. One pair was the duplicated Harker (2000) and Harker and Nash (1997) studies, which we report only once for obvious reasons. Within the Harker dataset, however, there were two different math comparisons available, year 9 and year 10. Although both are reported in Table 1, we included the second in our analysis as presumably this could represent more exposure to the gender composition of the school environment than would the first.

Several of the mathematics achievement outcomes reviewed by Mael et al. (2005) came from the High School and Beyond data set. We decided to use the comparison from Marsh (1991) because he reported *M*s and *SD*s for the uncontrolled comparisons, along with βs from the controlled analyses, and because there would be parallel effects available for the subsequent verbal and self-esteem/concept analyses reported here. This meant excluding Lee and Bryk (1986), Lee and Marks (1990), and Riordan (1990, 1994)---plus Marsh (1989), which was not listed for math, verbal or self-esteem by Mael et al. (2005), even though those variables are reported by Marsh (1989; see Appendix B in the online supplementary materials).

Rather than using the data from Daly and Shuttleworth (1997) for the three samples tested (Survey I, II, and III), as did Mael et al. (2005, Table 3, p. 19), we went back to the Daly (1996) results from girls in Survey I and Survey II, to avoid using results collapsed across gender from Daly and Shuttleworth (1997).

Finally, we did not include the Young and Fraser (1992) study in this comparison for two reasons. We do not see any evidence for the assertion by Mael et al. (2005, p. 23) that the DV in this study is math rather than physics as stated by Young and Fraser. In any case, there was not enough information to compute effect sizes.

This left us with 21 independent effect sizes (see Table 1) from nine studies. We found no significant effect of school gender composition, Hedges's g = 0.03, 95% CI = -0.03 to 0.09, p = .32. There was significant heterogeneity in effect sizes, $Q(20)$ = 445.9, p < .001, I^2 = 95.5, suggesting unexplained variance in the effect sizes. For this and subsequent analyses, we then tested for effects of the moderating variables we coded, such as gender, age, school type, etc. None of the potential moderators in this comparison was significant, however, including participant gender, $Q(2)$ = 3.57, p = .17.

Inspection of the control variables in Table 1 does suggest a relation between the size/direction of the effect size and the corresponding control. In other words, samples in which single-sex schools have better math achievement than do mixed-sex schools also tend to show a relation between higher SES or higher prior ability and a single-sex environment. There were 10 effect sizes for school gender composition and math achievement where there were also available effect sizes for school gender composition and a control variable of SES or prior ability. Correlating the school gender composition effects for math and control variables produced an association of borderline significance, $r(9) = 0.56$, $p = .09$.

Verbal achievement. Table 2 shows the studies with measures of verbal achievement, and the same information on uncontrolled and controlled comparisons, and control variables as shown in Table 1. Positive effect sizes indicate an advantage to single-sex schools, whereas negative effect sizes indicate an advantage to mixed-sex schools. The verbal achievement comparison is more problematic than was the one for math. First, there are several studies with multiple measures of verbal ability (reading, vocabulary, writing) but not enough information to compute averages within a sample. When a study reported multiple verbal measures, we used reading in the meta-analysis because it seemed to be the more general measure. Second, in this group of studies there are only 13 independent effect sizes. Overall, there was a significant association between school gender composition and verbal performance, Hedges's $g = 0.18$, 95% CI = 0.10 to 0.26. Again, there was unexplained heterogeneity in the effect sizes, $Q(11) = 28.8$, $p = .002$, $I^2 = 61.9$. None of the moderators was significant, including gender, $Q(1) = 0.49$, $p = 0.78$.

As with math, inspection of the control variables in Table 2 does suggest a relation between the size/direction of the effect size for the DV and for the control. For those nine effects for which we could also compute the control variable effect size, there was a large and significant association between the effect size for verbal achievement and the effect size for the control variable, $r(8) = 0.73$, $p = .03$.

Self-concept and self-esteem. Because the distinction made in the Mael et al. (2005) review between self-concept and self-esteem resulted in too few articles in each category to do a meta-analysis, we combined the two types of measures. Both are self-related measures that were being hypothesized to be positively impacted by single-sex education (Mael et al., 2005, p. 36). Table 3 displays these studies and effects.

Although we have been critical at times of Mael et al.'s (2005) summaries and conclusions, the dissertation by Cuddy (2000) in this section is very difficult to describe. Mael et al. (2005, p. 37) appear to interpret the multiple selves measure as indicating an advantage to girls in single-sex schools. The significant F value they cite for school gender composition (6.12) is on page 46 in Cuddy (2000), and is the effect for "Total Score" on the multiplicity measure. This score is described on page 45 as "based on the opposites and conflicts generated." Thus, we believe that for this score, higher is worse, and that therefore this effect shows an advantage to CE girls. Two other measures were used. One was the total percentage of false self-attributes, which showed no significant school gender composition effect but for which no means are reported, and the "Mean Liking" scores, for which a higher score means that a respondent rated a self-description as liked. Presumably this would mean that a higher score represented a higher self-esteem. We believe that this makes this last measure the one that is most like the others in this collection, and therefore the effects for Mean Liking are the ones from Cuddy (2000) used for the meta-analysis.

There were some other issues with the summaries and conclusions in Mael et al. (2005, Tables 11 & 12). In addition to the usual problem with HS&B overlapping data sets, there were also some discrepancies between the summaries and the overall conclusions in the tables that are noted in our Table 3 footnotes.

There were 18 effect sizes that could be computed (see Table 3). Overall, there was no significant association between school gender composition and self-esteem/self-concept, Hedges's g = -0.02, 95% CI = -0.12 to 0.08. The variability in this set of studies is again significant, $Q(17) = 39.8$, $p = .001$, $I^2 = 57.3$. Again, none of the moderators were significant, including gender, $Q(1) = 0.3$, $p = 0.61$.

A subset of the studies in this comparison again reported control information that could be converted to an effect size. Unlike math and verbal performance, self-esteem or self-concept effect sizes did not have a strong relation to the size/direction of the control variables, r (16) = .14, $p = .60$.

Control variables. The next analyses are aimed at addressing one of the key criticisms of the literature on single-sex versus mixed-sex school comparisons that there are pre-existing differences between the students in the two types of schools, and that those differences account for any outcome differences observed. The first choice in a meta-analysis would be to use the control variables in a standard moderator analysis, as was done above with gender and other moderators that were available for each study. Because the control variables were different across studies, and also not available or reported for every study, we did not have this option.

We chose instead to compare across the studies in Tables 1-3 the common control variables as a function of school gender composition. As noted earlier, the two most common control variables were SES and some sort of prior academic achievement.

For SES there are 19 independent effects for the relation with school gender composition, and there was a significant overall relation between higher SES and single-sex (rather than mixed-sex) schools, Hedges's $g = 0.17$, 95% CI = 0.10 to 0.25. Although this is not a large effect, it is roughly the same magnitude as the verbal effect size, and much larger than math or self-esteem/concept. There is also significant heterogeneity, $Q(18) = 55.8$, $p < .001$, $I^2 = 67.8$. None of the potential moderators was significant, including gender, $Q(2) =$

1.4, $p = .50$.

For the analysis on the control variable of prior academic achievement, we were able to examine math and other domains separately. This seemed important because of the gender stereotypic associations with mathematics achievement. Thus, the second analysis examined the seven effect sizes that represented prior math achievement in relation to school gender composition. This association was significant, Hedges's $g = 0.34$, 95% CI = 0.10 to 0.58, indicating that students in single-sex schools had higher math scores on entry than did students in mixed-sex schools. There is also significant heterogeneity, $Q(6) = 60.56$, $p < .001$, $I^2 = 90.1$, but not enough effects to pursue moderator comparisons.

The third analysis compared school gender composition to prior performance in one of these areas: verbal (including reading or vocabulary), general achievement, or IQ. For this set of 17 effects, there was no significant association, Hedges's $g = 0.03$, 95% CI = -0.06 to 0.12. There was also significant heterogeneity, $Q(16) = 213.8$, $p < .001$, $I^2 = 92.5$. Gender was the only moderator close to significance, and there were two categories (boys, both boys and girls) with small numbers but similar directions of effects. Combining boys and both boys and girls and comparing to girls only did produce a significant effect for gender, $Q(1) = 5.5$, $p = .019$. When participants were girls there was no relation between preexisting verbal differences and school gender composition, but for boys or samples with both genders, single-sex schools tended to have students with better incoming verbal performance.

Discussion

The results of the present meta-analytic critique showed that the methods of the Mael et al. (2005) vote-counting review of the single-sex education literature were seriously flawed in many ways. Additionally, our own meta-analyses of several of the main dependent variables in the single-sex education literature showed little to no effect of school gender composition on outcomes.

During the process of coding studies for our meta-analysis, it became clear that the original Mael et al. (2005) review had many fewer independent findings than implied from their summaries. The dependent findings used in their review included several articles by the same authors that appeared to report on the same data from the same sample (e.g., Caspi, 1995 and Caspi et al., 1993), as well as seven articles reporting comparisons from the High School and Beyond data set (see Appendix B in the online supplementary materials). Because Mael et al. used a vote counting method for drawing conclusions about the advantages of single-sex versus mixed-sex schools, the outcomes from these dependent data sets are over-represented in their review, drawing into question the validity of their conclusions.

Summary of Findings

Our meta-analytic critique of the Mael et al. (2005) review revealed small to non-existent differences in the outcomes of students at single-sex and coeducational schools. Of the three meta-analyses computed for the present study, two revealed no significant effect of school gender composition on outcomes. Specifically, in the domains of math and self-esteem/self-concepts, there was no advantage of either single-sex or mixed sex educational settings (*g*s = 0.03 and -0.02, respectively). The one domain that yielded an advantage for single-sex education was verbal achievement (g = .18). This is a weak effect, however, according to established guidelines for interpreting effect sizes (Cohen, 1969). Importantly, the results of our meta-analyses of the control variables reported for comparisons in each of these domains showed a significant relationship between school gender composition and student background variables, specifically socioeconomic status and prior math achievement for both boys and girls, and for boys, prior verbal achievement.

Relation to Other Works

The points outlined in our meta-analytic critique extend previous critiques of the single-sex education literature (e.g., Bracey, 2006; Smithers & Robinson, 2006) by systematically highlighting the flaws in both individual comparisons of single-sex and mixed-sex schools and the existing reviews of these studies. The results of our meta-analyses emphasize the importance of using quantitative review methods when drawing conclusions about a body of findings; narrative reviews and vote counting methods of review, such as those used by Mael et al. (2005), focus solely on the significance, rather than the magnitude, of group differences.

Limitations

The major weakness of our meta-analytic critique lies in its scope. We do not claim that these results are an exhaustive meta-analysis of the available literature comparing single-sex and coeducational environments, but rather that we have used meta-analysis as a quantitative tool to reassess the oft-cited Department of Education findings. Thus, there are obvious limitations to the generalizability of our results. Indeed, we believe these results highlight the need for a larger, comprehensive meta-analysis of the entire literature. Additionally, we did not incorporate newer studies published since the Mael et al. (2005) report. There are many new empirical studies that have been published in the United States since then because of the rise of public single-sex education, many of which appear in both volumes of the *Sex Roles* special issue (e.g., Hayes et al., 2011; Patterson &

Pahlke, 2011). Because we limited our sample of studies to those used in the Mael et al. (2005) report, our sample of independent effects in each of the major dependent variable categories was limited as well. Thus, the subsequent effect sizes calculated should be interpreted with this fact in mind.

Implications for Future Work

Comparisons between schools are, by nature, difficult to make, even when accounting for various levels of student and school factors. In our paper, we have highlighted the presence of many methodological shortcomings, especially the absence of control variables, in the particular case of single-sex versus coeducational school comparisons. The results of our meta-analytic critique have two important implications.

First, it is imperative that education researchers and policy-makers consider the nature and quality of the studies on which they base their agendas. The Mael et al. (2005) review is frequently cited in the debate over public single-sex education in the United States. Based on many of the problems inherent in the review that we uncovered in the process of our meta-analysis (and echoed by other scholars, e.g., Smithers & Robinson, 2006), we recommend citing the narrative summary of the Mael et al. results with caution.

Second, we believe these data have implications for researchers who attempt any type of cross-school comparisons, regardless of whether their focus is the efficacy of single-sex or any other type of education. When comparing the effects of two types of schools, the magnitude and direction of the effects cannot be understood without careful consideration of the inherent differences in the types of schools and the types of students they attract. Our analyses of the control variables reported in many of the studies comparing single-sex and mixed-sex environments highlight this point: there was a substantial relationship between enrollment in these different types of schools and students' pre-existing characteristics. In most cases, students enrolled in single-sex schools came from higher SES families and had higher levels of prior academic achievement than students enrolled in comparison mixed-sex schools. This is an important factor to consider in the implementation of single-sex education in the public sector in the United States; if many of the advantages of single-sex education in the private sector are tied to pre-existing differences in the students who elect to attend (and are themselves selected by schools), it may not be possible to generalize the results of these studies to predicting outcomes of public single-sex schools. However, preliminary studies of U.S. public single-sex schools suggest that many of these schools operate similarly to their private peers in terms of selectivity, and that comparisons of public single-sex and mixed sex schools must take into account these important student background factors as well (Hayes et al., 2011).

Our re-examination of the Mael et al. (2005) review highlights many of the problems inherent in the broader single-sex education literature. Research comparing single-sex to mixed-sex school environments has suffered from a variety of problems that include both theoretical and methodological flaws. The key problem, however, remains that there are systematic preexisting differences between students in the two types of schools, and that those differences are plausible explanations for any subsequent differences in outcomes observed. These problems are not new ones. Compounding this situation is a reliance on inadequate and misleading narrative summaries of the research, and the lack of a logical theoretical basis for many of the proposed effects. Together, these issues may be making it difficult to garner a wide consensus concerning the effects of school gender composition, but the examination of the research that is supposedly the best available suggests that the effect sizes are quite small and may be accounted for by prior student differences.

References

Australia Bureau of Statistics (1997). 4102.0- Australian Social Trends, 1997. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Previousproducts/80FBC4826C05C115CA2570EC001B193C?opendocument

Ainley, J., & Daly, P. (2002). Participation in science courses in the final year of high school in Australia: The influences of single-sex and coeducational schools. In A. Datnow & L. Hubbard (Eds.), *Gender in policy and practice: Perspectives on single-sex and coeducational schooling* (pp. 243–262). New York, NY: RoutledgeFalmer.

Baker, D. P., Riordan, C., & Schaub, M. (1995). The Effects of Sex-Grouped Schooling on Achievement: The Role of National Context. *Comparative Education Review*, *39*, 468–482. Retrieved from http://www.jstor.org/stable/1189147

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

Bornholt, L., & Möller, J. (2003). Attributions About Achievement and Intentions About Further Study in Social Context. *Social Psychology of Education, 6*, 217-231. doi:10.1023/A:1024715609124

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, West Sussex, UK: John Wiley & Sons, Ltd.

Bracey, G. W. (2006). *Separate but superior? A review of issues and data bearing on single-sex education.* (No. EPSL-0611-221-EPRU). Arizona State University: Education Policy Reseaerch Unit (EPRU), College of Education, Division of Educational Leadership and Policy Studies. Retrieved from http://nepc.colorado.edu/publication/separate-but-superior-a-review-issues-and-data-bearing-single-sex-education

Brutsaert, H. (2002). Pupils' perceptions of discipline and academic standards in Belgian coeducational and single-sex schools. *Evaluation and Research in Education*, 16, 71–81. doi:10.1080%2F09500790208667008

Brutsaert, H., & Bracke, P. (1994). Gender context of the elementary school: Sex differences in affective outcomes. *Educational Studies*, *20*(1), 3–11. doi:10.1080/0305569940200101

Bushman, B. J., & Wang, M. C. (2009). Vote-counting procedures in meta-analysis. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 207-220). New York, NY: Russell Sage.

Carpenter, P. W., & Hayden, M. (1987). Girls' academic achievements: Single-sex versus coeducational schools in Australia. *Sociology of Education*, *60*(3), 156–167. doi: 10.2307/2112273

Caspi, A. (1995). Puberty and the gender organization of schools: How biology and social context shape the adolescent experience. In L. J. Crockett & A. C. Crouter (Eds.), *Pathways through adolescence: Individual development in relation to social contexts* (pp. 57–74). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Caspi, A., Lynam, D., Moffitt, T. E., & Silva, P. A. (1993). Unraveling girls' delinquency: Biological, dispositional, and contextual contributions to adolescent misbehavior. *Developmental Psychology*, *29*(1), 19–30. doi:10.1037/0012-1649.29.1.19

Chadwell, D. (2010). *A gendered choice: Designing and implementing single-sex programs and schools*. Thousand Oaks, CA: Corwin.

Chute, E. (2012, February 5). Pittsburgh Westinghouse 6-12 school regroups after single-gender plan is scrapped. *Pittsburgh Post-Gazette*.

Cipriani-Sklar, R. (1997). A quantitative and qualitative examination of the influence of the normative and perceived school environments of a coeducational public school vs. a single-sex catholic school on ninth-grade girls' science self-concept and anxiety in the area of science education. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 57(10)*, 4312A. (UMI 9706808).

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic.

Conway, K. E. (1996). *Differential effects of single-sex education on the mathematical reasoning ability, verbal reasoning ability, and self-concept of high school girls.* (Unpublished Ed.D). Florida International University.

Cooke, J. (2005). A comparison of Japanese and American education systems, Part 1. *The Oregon Mathematics Teacher*, Jan/Feb 2005, 14-16. Retrieved from http://cimm.ucr.ac.cr/ciaem/articulos/evaluacion/internacionales/

Cuddy, A. R. (2000). *The development of self-concept in adolescent girls attending single-sex and coeducational schools: Exploring the influence of cognitive and social factors.* (Unpublished Ph.D.). Bryn Mawr College.

Daly, P. (1995). Science course participation and science achievement in single sex and co-educational schools. *Evaluation and Research in Education*, *9*, 91–98. doi: 10.1080/09500799509533376

Daly, P. (1996). The Effects of Single-Sex and Coeducational Secondary Schooling on Girls' Achievement. *Research Papers in Education: Policy and Practice*, *11*, 289–306. doi:

10.1080/0267152960110306

Daly, P., & Ainley, J. (1999). Student participation in mathematics courses in Australian secondary schools. *The Irish Journal of Education*, *30*, 77–95. Retrieved from http://www.jstor.org/stable/30077464

Daly, P., & Shuttleworth, I. (1997). Determinants of public examination entry and attainment in mathematics: Evidence on gender and gender-type of school from the 1980s and 1990s in Northern Ireland. *Evaluation and Research in Education*, *11*, 91–101. doi: 10.1080/09500799708666919

Dyer, G., & Tiggemann, M. (1996). The effect of school environment on body concerns in adolescent women. *Sex Roles*, *34*, 127–138. doi: 10.1007/BF01544800

Frasz, C., & Kato, K. (1999). The educational structure of the Japanese school system. Retrieved from http://www2.ed.gov/pubs/Research5/Japan/structure_j.html

Finn, J. D. (1980). Sex differences in educational outcomes: A cross-national study. *Sex Roles*, *6*, 9-26. doi:10.1007/BF00288358

Garcia, D. M. (1998). Single-sex versus coeducational public schooling for girls: A high school comparison study. *Dissertation Abstracts International, 59(07)*, 2434A. (UMI No. 9839064)

Gilroy, M. K. (1991). Single-sex schooling and its effect on achievement, attitudes and behaviors. *Dissertation Abstracts International, 51(07)*, 3611B. (UMI No. 9034425)

Granleese, J., & Joseph, S. (1993). Self-perception profile of adolescent girls at a single-sex and a mixed-sex school. *Journal of Genetic Psychology, 154,* 525-530. doi:10.1080/00221325.1993.9914750

Haag, P. (1998). Single-sex education in grades K-12: What does the research tell us? In S.

Morse (Ed.), *Separated by sex: A critical look at single-sex education for girls* (pp. 13-38). Washington, DC: AAUW Educational Foundation.

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*, 1706-1707. doi: 10.1126/science.1205031

Harker, R. (2000). Achievement, Gender, and the Single-Sex/Coed Debate. *British Journal of Sociology of Education*, *21*, 203–218. doi: 10.1080/713655349

Harker, R., & Nash, R. (1997). *School Type and the Education of Girls: Co-ed or Girls Only?* Paper presented at the annual meeting of the American Education Research Association, Chicago.

Harvey, T. J., & Stables, A. (1986). Gender differences in attitudes to science for third year pupils: An argument for single-sex teaching groups in mixed schools. *Research in Science & Technological Education, 4*, 163–170. doi: 10.1080/0263514860040206

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

Kessels, U., & Hannover, B. (2008). When being a girl matters less: Accessibility of gender-related self-knowledge in single-sex and co-educational classes and its impact on students' physics-related self-concept of ability. *British Journal of Educational Psychology, 78*, 273- 289. doi: 10.1348/000709907X215938

Lambert, J. (1998). An investigation of the difference in multidimensional self-concept between adolescent girls in single-sex and coeducational school settings. *Dissertation Abstracts International, 59(02)*, 876B. (UMI No. 9824795)

LePore, P. C., & Warren, J. R. (1997). A comparison of single-sex and coeducational Catholic secondary schooling: Evidence from the National Educational Longitudinal Study of 1988. *American Educational Research Journal*, *34*, 485–511. doi: 10.3102/00028312034003485

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

Lee, V. E., & Bryk, A. S. (1989). Effects of single-sex schools: Response to Marsh. *Journal of Educational Psychology, 81*, 647–650. doi:10.1037/0022-0663.81.4.647

Lee, V. E., & Lockheed, M. E. (1990). The Effects of Single-Sex Schooling on Achievement and Attitudes in Nigeria. *Comparative Education Review*, *34*, 209–231. Retrieved from http://www.jstor.org/stable/1187954

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

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Mael, F. A. (1998). Single-sex and coeducational schooling: Relationships to socioemotional and academic development. *Review of Educational Research, 68*, 101-129. doi:10.2307/1170751

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

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

Marsh, H. W. (1991). Public, Catholic single-sex, and Catholic coeducational high schools: Their effects on achievement, affect, and behaviors. *American Journal of Education*, *99*, 320–356. doi:10.1086/443985

Marsh, H. W., Smith, I. D., Marsh, M., & Owens, L. (1988). The transition from single-sex to coeducational high schools: Effects on multiple dimensions of self-concept and on academic achievement. *American Educational Research Journal*, *25*, 237–269. doi:10.2307/1163082

Mensinger, J. (2001). Conflicting gender role prescriptions and disordered eating in single-sex and coeducational school environments. *Gender and Education*, *13*, 417–429. doi:10.1080/09540250120081760

Mensinger, J. L. (2003). Disordered eating, adherence to the superwoman ideal, and gender socialization in coeducational and single-sex school environments. *Dissertation Abstracts International, 64*(03), 1552B.

Moore, M., Piper, V., & Schaefer, E. (1993*). Single-sex schooling and educational effectiveness: A research overview*. In D. K. Hollinger & R. Adamson (Eds.). *Single-sex schooling: Perspectives from practice and research* (pp. 7-68). Washington, DC: U.S. Department of Education.

NASSPE. (2012). *Single-Sex Schools / Schools with single-sex classrooms / what's the difference?* Retrieved from http://www.singlesexschools.org/schools-schools.htm

National Center for Education Statistics (n.d.). High School & Beyond (HS&B). Retrieved from http://nces.ed.gov/surveys/hsb/

Patterson, M. M., & Pahlke, E. (2011). Student characteristics associated with girls' success in a single-sex school. *Sex Roles, 65*, 737-750. doi: 10.1007/s11199-010-9904-1

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*, 175-181. doi: 10.1037/0021-9010.90.1.175

Proach, J. A. (2000). A study of the attitudes and academic achievement in biology of females in a single-sex school vs. a coeducational school in the Philadelphia Archdiocesan secondary schools. *Dissertation Abstracts International, 61(03)*, 937A. (UMI No. 9966004)

Riordan, C. (1985). Public and Catholic Schooling: The Effects of gender context policy. *American Journal of Education*, *93*, 518–540. Retreived from: http://www.jstor.org/stable/1085143

Riordan, C. (1990). *Girls and Boys in School: Together or Separate?* New York, NY: Teachers College Press.

Riordan, C. (1994). Single-gender schools: Outcomes for African and Hispanic Americans. *Research in Sociology of Education and Socialization*, *10*, 177–205.

Riordan, C. (2002). What do we know about the effects of single-sex schools in the private sector? Implications for public schools. In A. Datnow & L. Hubbard (Eds.), *Gender in Policy and Practice: Perspectives on Single-Sex and Coeducational Schooling* (pp. 10-30). New York: Routledge Falmer.

Robinson, P., & Smithers, A. (1999). Should the sexes be separated for secondary education? – Comparisons of single-sex and co-educational schools. *Research Papers* *in Education, 14*, 23-49. doi: 10.1080! /0267152990140103

Salomone, R. C. (2003). *Same, different, equal: Rethinking single-sex schooling*. New Haven, CT: Yale University Press.

Sanders, E. T. W. (1992). *Black Inner-City Males and the Milwaukee Public Schools Immersion Program: A Progress Report*. Paper presented at the annual meeting of the University Council for Educational Administration, Minneapolis, Minn.

Sanders, E. T. W., & Reed, P. L. (1995). An investigation of the possible effects of an immersion as compared to a traditional program for African American males. *Urban Education*, *30*, 93–112. doi: 10.1177/0042085995030001007

Sax, L. (2005). *Why gender matters*. New York; Doubleday.

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

Schneider, F. W., Coutts, L. M., & Starr, M. W. (1988). In favour of coeducation: The educational attitudes of students from coeducational and single-sex high schools. *Canadian Journal of Education*, *13*, 479-496. Retrieved from http://www.jstor.org/stable/1495288

Shadish, W. R., Robinson, L., & Lu, C. (1999). *ES: A computer program for effect size calculation*. St. Paul, MN: Assessment Systems Corporation.

Signorella, M. L., Frieze, I. H., & Hershey, S. W. (1996). Single-sex versus mixed-sex classes and gender schemata in children and adolescents. *Psychology of Women Quarterly, 20*, 599–607. doi:10.1111/j.1471-6402.1996.tb00325.x

Smithers, A., & Robinson, P. (2006). *The paradox of single-sex and co-educational schooling*.

Buckingham, England: Carmichael Press, University of Buckingham.

Spielhofer, T., O'Donnell, L., Benton, T., Schagen, S., & Schagen, I. (2002). *The impact of school size and single-sex education on performance.* (No. LGA Research Report 33). Slough, UK: NFER.

Spielhofer, T., Benton, T., & Schagen, S. (2004). A study of the effects of school size and single-sex education in English schools. *Research Papers in Education*, *19*, 133-159. doi: 10.1080/02671520410001695407

Stables, A. (1990). Differences between pupils from mixed and single-sex schools in their enjoyment of school subjects and in their attitudes to science and to school. *Educational Review*, *42*, 221–230. doi: 10.1080/0013191900420301

Steinback, M., & Gwizdala, J. (1995). Gender differences in mathematics attitudes of secondary students. *School Science and Mathematics*, *95*, 36–41. doi: 10.1111/j.1949-8594.1995.tb15721.x

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

Thompson, J. S. (2003). The Effect of Single-Sex Secondary Schooling on Women's Choice of College Major. *Sociological Perspectives*, *46*, 257–278. Retrieved from http://www.jstor.org/stable/1389762

Thompson, T., & Ungerleider, C. (2004). Single-sex schooling final report. Vancouver, British Columbia: Canadian Centre for Knowledge Mobilisation. Retrieved from http://www.cckm.ca/pdf/SSS%20Final%20Report.pdf

U.S. Department of Education. (2006). 34 C.F.R. § 106.34(b)(1).

Watson, C. M., Quatman, T., & Edler, E. (2002). Career aspirations of adolescent girls: Effects of achievement level, grade, and single-sex school environment. *Sex Roles*, *46*, 323–335. doi:10.1023/A:1020228613796

Wilson, D. B. (2010). A spreadsheet for calculating standardized mean difference type effect sizes. Retrieved from http://mason.gmu.edu/~dwilsonb/ma.html

Woodward, L. J., Fergusson, D. M., & Horwood, L. J. (1999). Effects of single-sex and coeducational secondary schooling on children's academic achievement. *Australian Journal of Education*, *43*, 142–156.

Young, D. J., & Fraser, B. J. (1992). *Sex Differences in Science Achievement: A Multilevel Analysis.* Paper presented at the annual meeting of the American Research Association, San Francisco, Calif.

Table 1
Mathematics Achievement and Associated Measured Control Effect Sizes as a Function of School Gender Composition

Study	DV	Subgroup	*n*	Mael conclusion	ES (LL/UL) without controls	ES (LL/UL) with control	Controls variable	Controls ES
Baker et al. (1995)	IEA SIMS	Belgium boys	~1357	CE was given lead with boys, SS with girls – even though the SS lead in girls was not as big. Effect sizes after controls were mislabeled as *ds* (were actually *bs*)	-0.06 (-0.27/0.05)	CE	language SES education goals school quality	N/A [a]
		Belgium girls	~1357		-0.05 (-0.16/0.06)	CE		
		Japan boys	~3802		-0.39 (-0.46/-0.33)	CE		
		Japan girls	~3803		-0.10 (-0.16/-0.04)	CE		
		New Zealand boys	~576		0.08 (-0.04/0.29)	SS		
		New Zealand girls	~576		0.13 (-0.08/0.24)	SS		
		Thailand boys	~1825		-0.37 (-0.46/-0.28)	CE		
		Thailand girls	~1826		0.63 (0.53/0.73)	SS		
Conway (1996)	SAT	Grade 12 girls	270	ND	0.13 (-0.11/0.37)	SS	Prior ability	0.45

Study	DV	Subgroup	*n*	Mael conclusion	ES (LL/UL) without controls	ES (LL/UL) with control	Controls variable	Controls ES
Daly (1996); Daly & Shuttleworth (1997) [b]	Public exam score	Survey I girls	798	Did not report these results	.05 (-0.09/ 0.19)	ND	IQ SES	0.26 -0.07
		Survey I boys & girls	1839	ND [c]	-0.05	ND	IQ SES	N/A -0.10
		Survey II girls	411	Did not report these results	0.05 (-0.15/ 0.24)	ND	IQ SES	0.27 0.28
		Survey II boys & girls	859	ND [c]	-0.001	CE	IQ SES	N/A 0.40
		Survey III boys & girls	1431	ND [c]	0.08 (-0.04/ 0.19)	ND	IQ SES	N/A 0.18
Harker & Nash (1997); Harker (2000)	4th form test	Year 9 girls	1831	ND [d]	0.11	CE	SES Prior ability	0.20 0.23
	School certificate test	Year 10 girls	1564	ND [d]	0.21 (0.10/ 0.31)	CE		

Study	DV	Subgroup	*n*	Mael conclusion	ES (LL/UL) without controls	ES (LL/UL) with control	Controls variable	Controls ES
Lee & Lockheed (1990)	Math test	boys	785	CE	0.11 (-0.04/ 0.25)	-0.22	Father occ	0.15
		girls	227	SS	0.66 (0.38/ 0.93)	0.24	Father occ	0.41
LePore & Warren (1997)	Math test	boys	159	SS	0.33 (0.01/ 0.65)	CE	SES Prior ability	0.03 0.38
		girls	140	ND	0.03 (-0.31/ 0.37)	SS	SES Prior ability	0.15 0.02
Marsh (1991) - HSB	Math test	boys & girls	972	ND	0.10 (-0.03/ 0.23)	0.10	SES Prior ability	0.06 0.07
Marsh et al. (1988)	School certificate test	boys	~490	ND	-0.10 (-0.27/ 0.08)		N/A [e]	
		girls	~510	ND	-0.08 (-0.25/ 0.09)			

Study	DV	Subgroup	*n*	Mael conclusion	ES (LL/UL)		Controls	
					without controls	with control	variable	ES
Riordan (1985)	Math test	boys	460	Difficult to compare;		0.11	SES	CE
		girls	439	Mael uses two		0.05	SES	SS
		boys & girls	899	groups that differ on more than one factor	0.11 (-0.02/ 0.24)		Various other factors	
Spielhofer (2002)	GCSE performance	Boys	approx. 180,000	ND	??	ND	N/A [f]	
		Girls	approx. 180,000	SS	??	SS		

Note. The studies listed in this table include those presented in Mael et al. Table 3 (2005, pp. 18-23) as "Mathematics Achievement Test Scores," plus mathematics effects from the same set of studies presented in Mael et al. but that were not listed in Mael et al. Table 3. The following abbreviations were used: ES = Hedges's *g*; LL/UL = 95% CI for *g*; ND = no significant difference reported but no effect size available nor indicator of direction of nonsignificant effect. CE = direction of effect shows an advantage for a mixed-sex (co-educational) school. SS = direction of effect shows an advantage for a single-sex school. Positive effect sizes indicate an advantage to SS schools; negative effects indicate an advantage to CE schools.

[a] Baker et al. (1995) report on a regression of single-sex attendance by a variety of control variables, but effect sizes cannot be computed.

[b] Daly (1995, 1996) and Daly and Shuttleworth (1997) Survey I and Survey II appear to be the same samples, each reporting on somewhat different dependent measures and different student groups.

[c] In Daly & Shuttleworth (1997) effect sizes could only be computed for boys and girls combined as the comparison was CE schools versus boys schools versus girls schools for all three surveys. These comparisons are apparently what was reported in Mael et al. Table 3 (2005, 19). Their overall conclusion is no difference, which is fair given the tiny effect sizes, but the text description is not clear.

[d] In Mael et al.'s Table 3, they do not seem to recognize that Harker & Nash (1997) and Harker (2000) are the same study. The two separate summaries correctly note that the apparent advantage to single-sex schools disappears when controls are applied, but only one of the two summaries correctly identifies the final analyses only being on female students, and that latter summary omits the results from one of the two math tests.
[e] No measured controls, but all students were from same area attending SS schools that transitioned to CE.
[f] Numerous controls, but no reports on specific comparisons.

Table 2

Verbal Achievement and Associated Measured Control Effect Sizes as a Function of School Gender Composition

Study	DV	Subgroup	*n*	Mael conclusion	ES (LL/UL)		Controls	
					without controls	with control	variable	ES
Conway (1996)	SAT	Grade 12 girls	270	ND	0.14 (-0.10/ 0.38)	SS	Prior ability	0.27
Daly (1996)	Verbal public exam score	Survey I girls	776	Did not report these results	0.01 (-0.13/ 0.16)	SS	IQ SES	0.26 -0.07
		Survey II girls	520	Did not report these results	0.15 (-0.02/ 0.33)	ND	IQ SES	0.27 0.28
Harker & Nash (1997); Harker (2000) [a]	4th form English test	Year 9 girls	104	ND [a]	0.17	CE	SES Prior ability	0.20 0.23
	School certificate English test	Year 10 girls	85	ND [a]	0.28 (0.18/ 0.37)	SS		
LePore & Warren (1997)	Reading test	boys	159	SS	0.30 (-0.02/ 0.61)	SS	SES Prior ability	0.03 0.30
		girls	140	ND	-.027 (-0.37/ 0.31)	CE	SES Prior ability	0.02 0.05
Marsh (1991) - HSB	Reading test	boys & girls	972	ND	0.10 (-0.03/ 0.23)	-0.02	SES Prior ability	0.06 0.21

Study	DV	Subgroup	*n*	Mael conclusion	ES (LL/UL)		Controls	
					without controls	with control	variable	ES
Marsh et al. (1988)	School certificate English test	boys	~490		0.07 (-0.11/ 0.25)		N/A [c]	
		girls	~510		0.00 (-0.17/ 0.17)			
Riordan (1985)	Reading test	boys	460	Difficult to compare;		0.05	SES	CE
		girls	439	Mael uses two		-0.06	SES	SS
		boys & girls	899	groups that differ on more than one factor	0.14 (0.01 / 0.27)		Various other factors	
Woodward (1999)	Word reading score	boys	324	SS	0.42 (0.19/ 0.65)			
		Girls	333	SS	0.48 (0.26/ 0.70)			
		Boys & girls	657			0.02	SES Prior ability	0.45 0.49

Note. The studies listed in this table include those presented in Mael et al. Table 5 (2005, pp. 27-31) as "Verbal and English Achievement Test Scores," plus verbal test effects from the same set of studies presented in Mael et al. but that were not listed in Mael et al. Table 5. The following abbreviations were used: ND = no significant difference reported but no effect size available nor indicator of direction of nonsignificant effect. CE = direction of effect shows an advantage for a mixed-sex (co-educational) school. SS = direction of effect shows an advantage for a single-sex school. If an asterisk or asterisks follow either

CE or SS, then there is some indicator that this difference was significant, but we could not compute an effect size. The absence of an asterisk therefore indicates that the effect was not reported to be significant. Positive effect sizes indicate an advantage to SS schools; negative effects indicate an advantage to CE schools.

[a] In Mael et al.'s Table 5 (2005, p. 28), they do not seem to recognize that Harker & Nash (1997) and Harker (2000) are the same study. The two separate summaries correctly note that the apparent advantage to single-sex schools disappears when controls are applied, but only one of the two summaries correctly identifies the final analyses only being on female students, and that latter summary omits the results from one of the two English tests.

[b] From Lee & Bryk (1986), Table 3, p. 386, 10th grade reading.

[c] No measured controls, but all students were from same area attending SS schools that transitioned to CE.

Table 3

Self-Esteem or Self-Concept and Associated Measured Control Effect Sizes as a Function of School Gender Composition

Study	DV	Subgroup	*n*	Mael conclusion	ES (LL/UL)		Controls	
					without controls	with control	variable	ES
Brutsaert & Bracke (1994)	self-esteem	Grade 6 boys	1130	SS	0.16 (0.04/0.28)	0.18	SES	-0.10
		Grade 6 girls	965	ND	-0.04 (-0.17/0.09)	0.10	SES	0.30
Cipriani-Sklar (1997)	general self-concept	Grade 9 girls	213	SS [a]	-0.02 (-0.28/ 0.25)		Urban area SES	ND ND
Conway (1996)	total self-concept	Grade 9 girls	293	Did not report these results	0.24 (0.06/ 0.42)	N/A	Prior ability	1.02
		Grade 12 girls	293	ND	0.06 (-0.17/ 0.29)	CE	Prior ability	0.27
Cuddy (2000)	multiple selves liking	Grade 6 girls	68	Did not report these results	0.31 (-0.16/ 0.78)	N/A [b]		
		Grade 9 girls	49	Did not report these results	0.37 (-0.19/ 0.93)	N/A [b]		
		Grade 12 girls	53	Did not report these results	-0.33 (-0.90/ 0.24)	N/A [b]		
							SES Prior ability	0.15 0.03

Study	DV	Subgroup	*n*	Mael conclusion	ES (LL/UL) without controls	ES (LL/UL) with control	Controls variable	Controls ES
Garcia (1998)	self-esteem	Grade 12 Black girls	105	ND	-0.54 (-0.93/-0.16)	-0.40	SES Prior ability	0.22 -1.22
		Grade 12 White girls	85	ND	0.04 (-0.38/ 0.47)	0.24	SES Prior ability	-0.58 -0.71
		Grade 12 Asian girls	47	ND	-0.27 (-0.09/0.30)	-0.11	SES Prior ability	-0.24 -1.60
Lambert (1998)	self-concept	Grade 12 girls	50	ND [c]	N/A	ND	SES Prior ability [d]	0.66 -0.23
LePore & Warren (1997)	self-esteem	boys	159	SS	0.17 (-0.14/0.49)	SS	SES Prior ability	0.03 0.30
		girls	140	ND	-0.06 (-0.40/0.28)	CE	SES Prior ability	0.15 0.40
Marsh (1991)	self-esteem	boys & girls	972	ND	-0.10 (-0.27/0.07)	0.10	SES Prior ability	0.06 0.21
Marsh et al. (1988)	self-concept	boys	~490	SS [e]	CE		N/A [f]	
		girls	~510	SS [e]	CE			

Study	DV	Subgroup	*n*	Mael conclusion	ES (LL/UL)		Controls	
					without controls	with control	variable	ES
Sanders (1992; Sanders & Reed, 1995)	self-esteem	Grade 3 boys	68	CE	-0.76 (-1.25/-0.26)	CE	Prior ability	0.65
		Grade 4 boys	49	CE	-0.24 (-0.81/0.32)	CE	Prior ability	0.10
		Grade 5 boys	35	CE	-0.06 (-0.71/0.59)	SS	Prior ability	-0.23

Note. The studies listed in this table include those presented in Mael et al. Tables 11 and 12 (2005, pp. 37-41) as "Self-Concept" and "Self-Esteem." The following abbreviations were used: ND = no significant difference reported but no effect size available nor indicator of direction of nonsignificant effect. CE = direction of effect shows an advantage for a mixed-sex (co-educational) school. SS = direction of effect shows an advantage for a single-sex school. If an asterisk or asterisks follow either CE or SS, then there is some indicator that this difference was significant, but we could not compute an effect size. The absence of an asterisk therefore indicates that the effect was not reported to be significant. Positive effect sizes indicate an advantage to SS schools; negative effects indicate an advantage to CE schools.

[a] In Mael et al.'s Table 11 (2005, p. 37), they also report on two other measures of self-concept used by Cipriani-Sklar (1997); thus, the Mael et al. conclusion is based on all three measures. We did not include these math and science self-concept measures as they were not widely enough used in other studies.

[b] Cuddy (2000) compared the SS and CE students on SES, reading vocabulary, and other measures. She concluded that the lack of differences on almost all measures meant the two groups were comparable.

[c] Mael et al. (2005, p. 37) did not check any boxes for an overall conclusion, but their summary clearly indicates a conclusion of no difference.

[d] Lambert (1998) reported both GPA and SAT. We chose the effect size for GPA because it included more areas of ability than the SAT score would.

[e] Mael et al. (2005, p. 38) marked advantage SS for both boys and girls, but the accompanying narrative summary clearly indicates a CE advantage.

[f] No measured controls, but all students were from same area attending SS schools that transitioned to CE.

Appendix A[1]

Research Used in Mael et al. (2005) Main Analysis with Key Sample Characteristics a

Variables Identified

[1] Appendices available from http://link.springer.com/article/10.1007/s11199-013-0

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Ainley & Daly (2002)	Australia	public, private, parochial	12th grade boys & girls (~7500)	not specified	physical science course participation- Table 14 other science course participation	Language Background SES earlier school achievement urban or nonurban environment school type
Baker et al. (1995)	1. Belgium 2. Japan 3. New Zealand 4. Thailand	unknown	1. 12th grade boys & girls (2714) 2. 12th grade boys & girls (7605) 3. 12th grade boys & girls (1152) 4. 12th grade boys & girls (3651)	not specified	12th grade math achievement- Table 3	language background SES educational aspirations school learning environment

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Bornholt & Möller (2003)	Australia	public	high school (ages 11-16) boys (706) high school (ages 11-16) girls (654)	born in Australia or emigrated from Asia	attributions for success in English-Table 14 attributions for failure in English- Table 14 attributions for success in math- Table 14 attributions for failure in math- Table 14 intention to participate in English intention to participate in math	matching for SES previous math and English participation neighboring urban areas all public schools
Brutsaert & Bracke (1994)	Belgium	parochial	6th grade boys (1130) 6th grade girls (965)	not specified	school commitment- Table 18 self-esteem- Table 12 sense of belonging in school- Table 18 study commitment- Table 18 mastery/environmental control stress fear of failure	teacher gender social class school size

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Carpenter & Hayden (1987)	1. Australia (Queensland) 2. Australia (Victoria)	public, private, parochial	1. 12th grade girls (460) 2. 12th grade girls (579)	not specified	2. Grade 12 grades – Table 2 (mislabeled) 2. Score on grade 12 public exam- Table 2 1. & 2. Science course-taking	Father occupation Father education Mother education Teacher encouragement Parent encouragement Peer plans
1. Caspi et al. (1993) 2. Caspi (1995)	New Zealand	unknown	girls tested at ages 13, 15, 16 (297)	mostly white	1. & 2. Norm-breaking behavior (age 13)- Table 13 1. & 2. Self-reported delinquency (age 15)- Table 17 1. & 2. Familiarity with delinquent peers (age 13)- Table 17 2. School certificate exam (age 15) – Table 2 2. Sixth form exam (age 16) – Table 2	1. & 2. Timing of menarche 1. & 2. parental values 1. & 2. SES 1. & 2. childhood behavior problems 2. IQ

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Cipriani-Sklar (1997)	U.S.	mixed sex: public single-sex: parochial	9th grade girls (213)	white, black, Hispanic, Asian, Pacific Islander	general self-concept-Table 11 math self-concept-Table 11 science self-concept-Table 11 science state anxiety-Table 11 science trait anxiety-Table 11	same urban area
Conway (1996)	U.S.	parochial	9th grade girls (455-471) 12th grade girls (270-387)	white, Hispanic	SAT M - 12th grade-Table 3 SAT V - 12th grade-Table 5 self-concept - 9th grade- Table 12 self-concept 12th grade- Table 12	HS placement test verbal (9th & 12th graders) HS placement test quantitative (9th & 12th graders)

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Cuddy (2000)	U.S.	private	6th grade girls (68) 9th grade girls (49) 12th grade girls (53)	white, black, Hispanic, Asian, other	Harter multiple selves total score- Table 11 School climate –Table 31 (mislabeled as Harter multiple selves false self score) Harter multiple selves liking score Harter multiple selves false self score Rating of characteristics leading to popularity in school Harter SPPC (3 subscales) Harter SPPA (4 subscales)	Vocabulary test SES parent education

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
1. Daly (1995) 2. Daly (1996) 3. Daly & Shuttleworth (1997) Survey I	U.K. (Northern Ireland)	unknown	1. 16-year-old boys & girls (1234-2523) 2. 16-year-old girls (798-1077) 3. 16 year old boys & girls (1839-2566) all 3 apparently the same sample	unknown	Total examination score – Table 2 science course enrollment- Table 14 science achievement (public exams physics, chem, bio)- Table 4 Math exam entry- Table 14 Math achievement- Table3 English achievement	verbal reasoning score community religion at grammar school SES
1. Daly (1995) 2. Daly (1996) 3. Daly & Shuttleworth (1997) Survey II	U.K. (Northern Ireland)	unknown	1. 16-year old boys & girls (603-1584) 2. 16 year old girls (411-782) 3. 16 year old boys & girls (859-1585) all 3 apparently the same sample	unknown	Total examination score- Table 2 science course enrollment- Table 14 science achievement (public exams physics, chem, bio)- Table 4 Math entry exam- Table 14 Math achievement- Table 3 English achievement	IQ community religion at grammar school SES

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Daly & Shuttleworth (1997) Survey III	U.K. (Northern Ireland)	unknown	16 year old boys and girls (1286-1431)	unknown	Math achievement- Table 3 Math entry exam- Table 14	community religion at grammar school SES
Daly, Ainley, & Robinson (1996, cited in Mael et al., 2005) - could not retrieve						
Daly & Ainley (1999)	Australia	public, private, parochial	12th grade boys (798) 12th grade girls (1204)	English speaking background or not	participation in types of math courses (advanced, general, none)- Table 14	earlier math performance parent education language background type of school urban or not

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Garcia (1998)	U.S.	public	12th grade girls (243)	white, black, Asian	SAT Total- Table 2 self-esteem- Table 12 leadership opportunities – Table 30 GPA anticipated college enrollment academic aspirations professional aspirations personal aspirations	SES previous education 9th grade entrance exam score
Harker (2000); Harker & Nash (1997) - same data reported in both	New Zealand	public	boys & girls in years 9, 10, & 11 in school (ns vary and major comparisons are on girls only)	white, Asian, Maori, Pacific Islander	Fourth form (year 9) math, girls only- Table 3 School certificate (Yr. 10) math, girls only- Table 3	initial ability SES ethnicity

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
High School and Beyond dataset (7 studies)	U.S.	public, private, parochial most articles focus on the Catholic subsample to provide a more controlled comparison	boys & girls tested as sophomore, seniors, and post-high school	diverse	many--see Table 2 for details	many--see Table 2 for details
Lambert (1998)	U.S.	parochial	12th grade girls (50)	not specified	self-concept (7 domains)- Table 11	SES GPA SAT

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Lee & Lockheed (1990)	Nigeria	public	15-16 year old boys (785) 15-16 year old girls (227)	not specified	Math achievement- Table 3 Math stereotyping- Table 14	age father occupation rural or not home language educational expectation perceived self ability parental support motivation school size length school year student/teach ratio teacher gender teacher preparation (2 vars) Instructional quality (7 vars)
LePore & Warren (1997)	U.S.	parochial	boys in grades 8, 10, & 12 (159) girls in grades 8, 10, 12 (140) (both of these are the final, weighted sample size)	white, non-white	Math- Table 3 Science- Table 4 Reading- Table 5 self-esteem- Table 12 locus of control- Table 13 Social studies	SES number of siblings % two-parent household race/ethnicity educational aspirations grade 8 test scores

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Marsh et al. (1988)	Australia	public	boys & girls in grades 8-12 and over time; ns vary based on the specific comparison	not specified	Math exam (Gr. 10)- Table 3 English exam (Gr. 10)- Table 5 self-concept- Table 11	pre- post-test design
Mensinger (2001) (reanalysis of Dyer & Tiggemann, 1996)	Australia	private	girls ages 15-17 (142)	white	EDI subscales Drive for Thinness, Bulimia, Body dissatisfaction- Table 24	school SES
Proach (2000)	U.S.	parochial	girls 10^{th} grade (48)	diverse	biology test – Table 4 gender equity in science science attitudes perception of science and scientists	pre- post-test design ethnicity

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Riordan (1985); Riordan (1990)	U.S.	mixed sex: public, parochial single-sex: parochial	1972 12th grade boys (4873) 12th grade girls (5552) 1979 follow up men (~25-250) women (~25-250) 1986 follow up men (~75-100) women (~75-100)	white	1972 data: Advanced Math-Table 3 Science – Table 4 Civics- Table 7 Self-esteem- Table 11 locus of control- Table 13 attitudes toward working women- Table 20 1979 outcomes: verbal- Table 8 math- Table 8 1986 outcomes: marital happiness Writing educational attainment occupation self-esteem locus of control attitudes toward women's roles	SES region of U.S. 1979: 1972 verbal + math

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Sanders (1992) (published in Sanders & Reed, 1995)	U.S.	public	3rd grade boys (69) 4th grade boys (49) 5th grade boys (35)	black	Self-esteem -Table 12 Attitude toward school - Table 14 Intellectual Achievement Responsibility Questionnaire Table 18	vocabulary reading math previous year attendance
Schneider & Coutts (1982) (same sample as Schneider et al., 1988)	Canada	parochial	10th grade boys (504) 10th grade girls (510) 12th grade boys (506) 12th grade girls (509)	unknown	value climate (6 items) - Table 29 "environmental press": - Table 29 Strong intellectual orientation School activities Strong environmental control Negative attitude	SES

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Spielhofer et al. (2002) (published in Speilhofer et al., 2004)	U.K. (England)	public, private, parochial	year 11 boys & girls (369,341)	unknown	Total GCSE score - Table 2 Average GCSE points score - Table 2 Number of GSCEs taken Table 2 Number of science GSCEs taken - Table 14 entered for higher math - Table 14 Modern Languages taken - Table 14 Food Technology taken - Table 14 Resistant Materials taken - Table 14 Math score English score Total science score Average science score Graphics	previous math level previous English level previous science level age selectivity of school free school meal % class size school size presence and size of 6th form

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Stables (1990) (some data also appear to be in Harvey & Stables, 1986)	U.K. (England)	public	13-14 year old boys and girls (>2300)	unknown	Science attitudes (all science and by subarea) - Table 14 Subject preferences (all subjects) Perception of subject importance (all subjects)	verbal reasoning quotient
Steinback & Gwizdala (1995)	U.S.	parochial	girls apparently in grades 9-12 353 in year 1, 323 in year 2	unknown	Math attitudes (14 questions examined separately) - Table 14	pre- post-test design
Watson et al. (2002)	U.S.	mixed sex: public single-sex: unknown	girls in grade 10 (294) girls in grade 12 (200)	white, black, Asian, Hispanic, other	ideal job choice - Table 16 realistic job choice - Table 16	SES teacher ratings of student achievement

Study	Country	School Type	Gender and Age (n)	Ethnicity	Dependent variables	Control variables
Woodward et al. (1999)	New Zealand	public, private, parochial	boys tested at various times before, during, and after high school (324) girls tested at various times before, during, and after high school (333)	unknown	school certificate attainment - Table 2 reading - Table 5 entry to seventh form leaving school before age 17 - Table 21 leaving school w/o qualifications unemployment Table 23	age 13 Test of Scholastic ability age 10 & 12 reading comprehension age 11 math reasoning age 8 IQ age 8 behavior problems maternal age maternal education SES maternal discipline maternal emotional responsiveness parent change parent conflict number of schools attended source of school funding
Young & Fraser (1992)	Australia	public, private, parochial	14 year old boys (2352) 14 year old girls (2565)	unknown	Physics achievement - Table 3 Math (Mael reviewers insisted on putting this in the math section)	SES School type Student attitude Verbal & Math ability

Note. The studies listed in this table are referred to in Mael et al. (2005) as "Quantitative Phase III-Coded references" (pp. 94-97). Additional details on the seven High School and Beyond studies are in Table 2. All study details in this table are a result of the coding by the present authors. Dependent variables with a table notation were used in the Mael et al. (2005). Dependent variables at the bottom of a study cell with no table notation were variables in the study that were not used in Mael et al. (2005).

Appendix B[2]

Articles in Mael et al. (2005) using the High School and Beyond (HS&B) Data Set wit Dependent and Control Variables Identified

[2] Appendices available from http://link.springer.com/article/10.1007/s11199-013-0

Study	Dependent Variables	Control Variables
Lee & Bryk (1986)	Math achievement- Table 3 Science achievement- Table 4 Writing achievement- Table 5 Reading achievement- Table 5 Self-concept- Table 11 Locus of control- Table 13 Math courses (years)- Table 14 Math stereotyping- Table 14 Educational aspirations- Table 15 Interest in math- Table 14 & 18 Physical Science courses (years) – Table 14 Vocational courses (yrs)- Table 14 Interest in English- Table 14 & 18 Academic friends- Table 14 & 18 Discipline problems- Table 17 Unexcused absences- Table 17 Homework (hours/week)- Table 19 Sex stereotyping- Table 26 Attitudes towards social peers Attitudes towards student athletes TV (hours/week) Social studies courses (years)	Academic track Social class Race Single parent Religion Elementary school Grade repeat Grade 8 college plans Financial sacrifice

Study	Dependent Variables	Control Variables
Lee & Marks (1990)	SAT M- Table 3 SAT V- Table 5 Educational Aspirations, 2 years post HS- Table 15, 16 Sex stereotype of work (4 years post HS)- Table 26 Political involvement/activism- Table 27 College satisfaction, academic- Table 33 College satisfaction, social- Table 33 Reading achievement- Table 5 Math achievement- Table 3 Science achievement- Table 4 Locus of control- Table 13 Self-esteem- Table 12 Type of college Self-concept (4 yrs post HS) Locus of control (4 yrs post HS) Sex stereotyping of family (4 yrs post HS) Political involvement, disc Work orientation, humanistic Work orientation, prestige Financial reward orientation Social justice orientation Social studies achievement	Academic track Social class Race Single parent Religion Elementary school Grade repeat Grade 8 college plans Financial sacrifice

Study	Dependent Variables	Control Variables
Marsh (1989)	Math course pattern- Table 14 Math credits- Table 14 Science course pattern- Table 14 Physical science credits- Table 14 Life science credits- Table 14 Vocational course pattern- Table 14 Foreign language credits- Table 14 Social science credits- Table 14 English credits- Table 14 Academic credits- Table 14 Academic track- Table 14 Vocational track- Table 14 Honors- Table 14 Math achievement Reading achievement Science achievement Writing achievement Vocabulary achievement Grades Self-esteem Work Community Locus of control Family Academic self-concept Social self-concept Appearance self-concept Athletic self-concept Sex stereotypes Educational Aspirations Parent aspirations Parent involvement	Individual level variables: SES Mother employment status Race: Black Race: Hispanic Single parent Non-Catholic Grade repeat College expectations All Catholic schools All public schools Kindergarten attendance Home language: English Family size Religion School level variables: Year enrollment Mean SES Mean % Black Mean % Hispanic Urbanicity Community income Mean achievement test score

Study	Dependent Variables	Control Variables
Marsh (1991)	Math- Table 3 Science- Table 4 Reading- Table 5 Writing- Table 5 Vocabulary- Table 5 Grades- Table 6 Self-concept- Table 11 Locus of control- Table 13 Math course homework pattern- Table 14 Science course homework pattern- Table 14 Vocational course homework pattern- Table 14 Academic credits- Table 14 Academic track- Table 14 Vocational track- Table 14 Honors- Table 14 Homework- Table 19 Educational aspirations- Table 15 Trouble at school- Table 17 Postsecondary education- Table 22 Unemployed- Table 23 Academic self-concept Parental investment School discipline policy	SES Race Gender Grade repeat College expectations

Study	Dependent Variables	Control Variables
Riordan (1990, ch. 5)	Advanced math- Table 3 Science- Table 4 Civics- Table 7 Self-esteem - Table 11 Locus of control- Table 13 Attitudes towards working women- Table 20 Writing	Race SES Parent income Initial ability Track Relevant course work Amount of Homework Dating Attitudes towards good students Attitudes towards athletes
Riordan (1994)	Test score composite- Table 2 Self-esteem- Table 12 Locus of control- Table 13 Leadership composite- Table 30	SES Parent income Family size Family structure Mother work status Region Track Coursework Homework Parental interest Youth culture Level of discipline

Study	Dependent Variables	Control Variables
Thompson (2003)	College major (Masc., Neutral, Fem.) - Table 25	SES single parent religiosity years in Catholic elementary school school ethnicity 4 year college English test scores (gr. 10 & 12) math test scores (gr. 10 & 12) science test scores (gr. 10 & 12) feminists attitudes (gr. 10 & 12) math courses science courses

Note. The studies listed in this table are referred to in Mael et al. (2005) as "Quantitative Phase III-Coded references" (pp. 94-97), but are all based on the same data set (High School and Beyond [HSB]). All study details in this table are a result of the coding by the present authors. Dependent variables with a table notation were used in the Mael et al. (2005) review. Dependent variables at the bottom of a study cell with no table notation were variables in the study that were not used in Mael et al. (2005).

References

Ainley, J., & Daly, P. (2002). Participation in science courses in the final year of high school in Australia: The influences of single-sex and coeducational schools. In A. Datnow & L. Hubbard (Eds.), *Gender in policy and practice: Perspectives on single-sex and coeducational schooling* (pp. 243–262). New York, NY: RoutledgeFalmer.

Baker, D. P., Riordan, C., & Schaub, M. (1995). The Effects of Sex-Grouped Schooling on Achievement: The Role of National Context. *Comparative Education Review*, *39*, 468–482. Retrieved from http://www.jstor.org/stable/1189147

Bornholt, L., & Möller, J. (2003). Attributions About Achievement and Intentions About Further Study in Social Context. *Social Psychology of Education, 6*, 217-231. doi:10.1023/A:1024715609124

Brutsaert, H., & Bracke, P. (1994). Gender context of the elementary school: Sex differences in affective outcomes. *Educational Studies*, *20*(1), 3–11. doi:10.1080/0305569940200101

Carpenter, P. W., & Hayden, M. (1987). Girls' academic achievements: Single-sex versus coeducational schools in Australia. *Sociology of Education*, *60*(3), 156–167. doi: 10.2307/2112273

Caspi, A. (1995). Puberty and the gender organization of schools: How biology and social context shape the adolescent experience. In L. J. Crockett & A. C. Crouter (Eds.), *Pathways through adolescence: Individual development in relation to social contexts* (pp. 57–74). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Caspi, A., Lynam, D., Moffitt, T. E., & Silva, P. A. (1993). Unraveling girls' delinquency: Biological, dispositional, and contextual contributions to adolescent misbehavior.

Developmental Psychology, *29*(1), 19–30. doi:10.1037/0012-1649.29.1.19

Cipriani-Sklar, R. (1997). A quantitative and qualitative examination of the influence of the normative and perceived school environments of a coeducational public school vs. a single-sex catholic school on ninth-grade girls' science self-concept and anxiety in the area of science education. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 57(10)*, 4312A. (UMI 9706808).

Conway, K. E. (1996). *Differential effects of single-sex education on the mathematical reasoning ability, verbal reasoning ability, and self-concept of high school girls.* (Unpublished Ed.D). Florida International University.

Cuddy, A. R. (2000). *The development of self-concept in adolescent girls attending single-sex and coeducational schools: Exploring the influence of cognitive and social factors.* (Unpublished Ph.D.). Bryn Mawr College.

Daly, P. (1995). Science course participation and science achievement in single sex and co-educational schools. *Evaluation and Research in Education*, *9*, 91–98. doi: 10.1080/09500799509533376

Daly, P. (1996). The Effects of Single-Sex and Coeducational Secondary Schooling on Girls' Achievement. *Research Papers in Education: Policy and Practice*, *11*, 289–306. doi: 10.1080/0267152960110306

Daly, P., & Ainley, J. (1999). Student participation in mathematics courses in Australian secondary schools. *The Irish Journal of Education*, *30*, 77–95. Retrieved from http://www.jstor.org/stable/30077464

Daly, P., & Shuttleworth, I. (1997). Determinants of public examination entry and attainment in mathematics: Evidence on gender and gender-type of school from the

1980s and 1990s in Northern Ireland. *Evaluation and Research in Education*, *11*, 91–101. doi: 10.1080/09500799708666919

Dyer, G., & Tiggemann, M. (1996). The effect of school environment on body concerns in adolescent women. *Sex Roles*, *34*, 127–138. doi: 10.1007/BF01544800

Garcia, D. M. (1998). Single-sex versus coeducational public schooling for girls: A high school comparison study*. Dissertation Abstracts International, 59(07)*, 2434A. (UMI No. 9839064)

Harker, R. (2000). Achievement, Gender, and the Single-Sex/Coed Debate. *British Journal of Sociology of Education*, *21*, 203–218. doi: 10.1080/713655349

Harker, R., & Nash, R. (1997). *School Type and the Education of Girls: Co-ed or Girls Only?* Paper presented at the annual meeting of the American Education Research Association, Chicago.

Harvey, T. J., & Stables, A. (1986). Gender differences in attitudes to science for third year pupils: An argument for single-sex teaching groups in mixed schools. *Research in Science & Technological Education, 4*, 163–170. doi: 10.1080/0263514860040206

Lambert, J. (1998). An investigation of the difference in multidimensional self-concept between adolescent girls in single-sex and coeducational school settings*. Dissertation Abstracts International, 59(02)*, 876B. (UMI No. 9824795)

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

Lee, V. E., & Lockheed, M. E. (1990). The Effects of Single-Sex Schooling on Achievement and Attitudes in Nigeria. *Comparative Education Review*, *34*, 209–231. Retrieved from http://www.jstor.org/stable/1187954

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

LePore, P. C., & Warren, J. R. (1997). A comparison of single-sex and coeducational Catholic secondary schooling: Evidence from the National Educational Longitudinal Study of 1988. *American Educational Research Journal*, *34*, 485–511. doi: 10.3102/00028312034003485

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

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

Marsh, H. W. (1991). Public, Catholic single-sex, and Catholic coeducational high schools: Their effects on achievement, affect, and behaviors. *American Journal of Education*, *99*, 320–356. doi:10.1086/443985

Marsh, H. W., Smith, I. D., Marsh, M., & Owens, L. (1988). The transition from single-sex to coeducational high schools: Effects on multiple dimensions of self-concept and on academic achievement. *American Educational Research Journal*, *25*, 237–269. doi:10.2307/1163082

Mensinger, J. (2001). Conflicting gender role prescriptions and disordered eating in single-sex and coeducational school environments. *Gender and Education*, *13*, 417–429. doi:10.1080/09540250120081760

Proach, J. A. (2000). A study of the attitudes and academic achievement in biology of females in a single-sex school vs. a coeducational school in the Philadelphia Archdiocesan secondary schools. *Dissertation Abstracts International, 61(03)*, 937A. (UMI No. 9966004)

Riordan, C. (1985). Public and Catholic Schooling: The Effects of gender context policy. *American Journal of Education*, *93*, 518–540. Retreived from: http://www.jstor.org/stable/1085143

Riordan, C. (1990). *Girls and Boys in School: Together or Separate?* New York, NY: Teachers College Press.

Riordan, C. (1994). Single-gender schools: Outcomes for African and Hispanic Americans. *Research in Sociology of Education and Socialization*, *10*, 177–205.

Sanders, E. T. W. (1992). *Black Inner-City Males and the Milwaukee Public Schools Immersion Program: A Progress Report*. Paper presented at the annual meeting of the University Council for Educational Administration, Minneapolis, Minn.

Sanders, E. T. W., & Reed, P. L. (1995). An investigation of the possible effects of an immersion as compared to a traditional program for African American males. *Urban Education*, *30*, 93–112. doi: 10.1177/0042085995030001007

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

Spielhofer, T., O'Donnell, L., Benton, T., Schagen, S., & Schagen, I. (2002). *The impact of school size and single-sex education on performance.* (No. LGA Research Report 33). Slough, UK: NFER.

Spielhofer, T., Benton, T., & Schagen, S. (2004). A study of the effects of school size and single-sex education in English schools. *Research Papers in Education*, *19*, 133-159. doi: 10.1080/02671520410001695407

Stables, A. (1990). Differences between pupils from mixed and single-sex schools in their enjoyment of school subjects and in their attitudes to science and to school. *Educational Review*, *42*, 221–230. doi: 10.1080/0013191900420301

Steinback, M., & Gwizdala, J. (1995). Gender differences in mathematics attitudes of secondary students. *School Science and Mathematics*, *95*, 36–41. doi: 10.1111/j.1949-8594.1995.tb15721.x

Thompson, J. S. (2003). The Effect of Single-Sex Secondary Schooling on Women's Choice of College Major. *Sociological Perspectives*, *46*, 257–278. Retrieved from http://www.jstor.org/stable/1389762

Watson, C. M., Quatman, T., & Edler, E. (2002). Career aspirations of adolescent girls: Effects of achievement level, grade, and single-sex school environment. *Sex Roles*, *46*, 323–335. doi:10.1023/A:1020228613796

Woodward, L. J., Fergusson, D. M., & Horwood, L. J. (1999). Effects of single-sex and coeducational secondary schooling on children's academic achievement. *Australian Journal of Education*, *43*, 142–156.

Young, D. J., & Fraser, B. J. (1992). *Sex Differences in Science Achievement: A Multilevel Analysis.* Paper presented at the annual meeting of the American Research

Association, San Francisco, Calif.