DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 48, ARTICLE 33, PAGES 931–944
PUBLISHED 21 JUNE 2023
https://www.demographic-research.org/Volumes/Vol48/33/
DOI: 10.4054/DemRes.2023.48.33

Descriptive Finding

The gender gap in schooling outcomes: A cohort study of young men and women in India

Nabamita Dutta

John Nunley

Contents

Demographic Research: Volume 48, Article 33
Descriptive Finding

The gender gap in schooling outcomes: A cohort study of young men and women in India

Nabamita Dutta[1]

John Nunley[2]

Abstract

BACKGROUND
Although literacy rates in India have improved for both men and women, less is known about the evolution of gender disparities across different levels of educational attainment.

OBJECTIVE
The goal is to determine whether gender gaps in schooling outcomes have narrowed, widened, or remained unchanged across birth cohorts.

METHODS
With a multinomial logit specification, we compare six education outcomes for people born to the 1956–1960, 1961–1965, …, and 1986–1990 cohort groups. Our empirical tests indicate whether the gender gaps have narrowed, widened, or remained unchanged across the cohort groups.

RESULTS
We find evidence of narrowing gender gaps for some but not all education outcomes. The gender gaps narrow for not attending school, attending primary school, and primary school completion, but they persist for secondary school completion, attending college, and college completion.

CONCLUSIONS
Although we observe improvements in the gender gaps in schooling outcomes toward the lower end of the education spectrum, gender inequities associated with higher levels of schooling persist across cohort groups. It is important to understand the causes of these patterns, as there are likely important policy considerations for India as it grapples with the interactions among technological change, a relatively young workforce, and persistent gendered norms and attitudes.

[1] Department of Economics, University of Wisconsin–La Crosse, USA. Email: ndutta@uwlax.edu.
[2] Department of Economics, University of Wisconsin–La Crosse, USA. Email: jnunley@uwlax.edu.

CONTRIBUTION

The paper makes two noteworthy contributions. First, we show that gender progress in schooling outcomes is not uniform across different levels of educational attainment. Second, our cohort study framework provides a simple test for progress (or lack thereof) in education and other settings.

1. Introduction

Despite economic development rankings that are similar to those for Brazil, Russia, and China, India fares worse in terms of educational attainment (Kingdon 2007). Studies such as Dhar, Jain, and Jayachandran (2018) find that boys and girls begin secondary schooling at the same rate, but for every boy enrolled in tertiary education, only around 0.7 girls are enrolled. The World Economic Forum (2020) ranks India 112th out of 153 countries in its Global Gender Gap Index. However, literacy rates have improved substantially in recent decades, and growth in literacy rates has disproportionately benefited women as the gender gap in literacy rates has declined over the last 20 years (Katiyar 2016).

Despite the lowering of the gender gap in literacy rates, less is known about how gender disparities have evolved for different levels of educational attainment. Although studies vary in terms of magnitude, the returns on schooling in India are large, and they are larger for women (Duraisamy 2002; Mitra 2019). Moreover, higher levels of educational attainment for women tend to stimulate economic activity (Self and Grabowski 2004). Thus, from a policy standpoint, increasing educational attainment, particularly for women, is a promising policy lever with which to improve general well-being and expand economic output in general.

In this paper, we ask whether women are catching up or falling further behind men in terms of six educational outcomes: not attending school, attending some primary school but not completing it, completing primary school, attending some secondary school, completing secondary school/attending some college, and college completion. Given that levels of educational attainment for women have implications for their empowerment, employment, access to financial services, and self-employment, we aim to fill a gap in the literature by evaluating progress (or lack thereof) in education outcomes for India. Using a multinomial logit specification, we propose a simple test that relies on a cohort study framework that holds age effects constant, allowing us to avoid conflating cohort effects with age effects. We measure whether the gender gap in education outcomes has narrowed, widened, or remained unchanged across the 1956–1960, 1961–1965, …, 1981–1985, and 1986–1990 birth cohort groups. Given that annual data over a lengthy time horizon are unavailable for India, our analysis pools cross-sectional data

collected at different points in time from India's Socio-Economic Survey (SES). These data span 26 years, as the earliest survey available from IPUMS International is from 1983 and the most recent is from 2009 (Minnesota Population Center 2020).

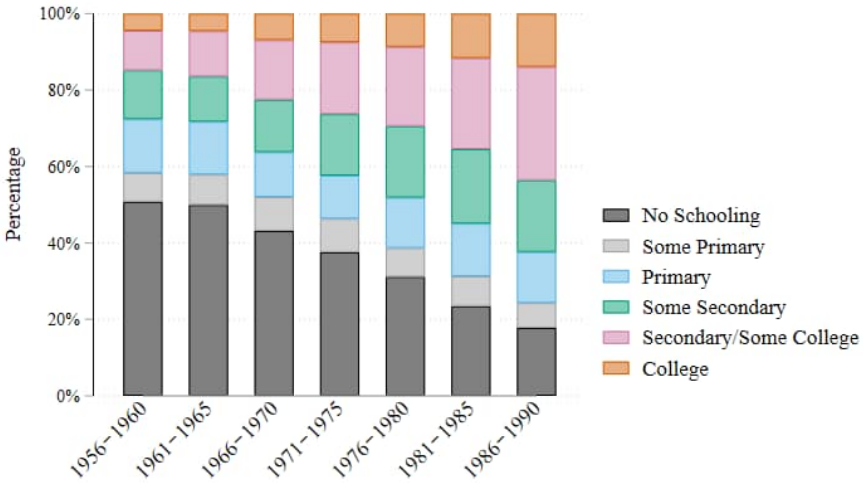
We first show the dramatic improvements in all schooling outcomes for India as a whole and that these gains are, at least to some degree, shared among men and women. Using a multinomial specification, we compute adjusted risk ratios (ARRs) and adjusted risk differences (ARDs), following the approach advanced by Norton, Miller, and Kleinman (2013). The estimated ARRs and ARDs are informative, but the interpretation of the ARRs and ARDs is inconsistent for particular education outcomes (e.g., college completion). Consider the no-schooling rate. Among men, those born to the most recent cohort (1986–1990) are about 72% more likely to complete at least some schooling than their counterparts born to the earliest cohort (1956–1960). The same calculation for women yields a 65% greater likelihood of completing any amount of schooling. A comparison of these numbers suggests that the rate at which men move from no schooling to some schooling exceeds that of women, which is suggestive of a widening of the gender gap for this particular educational outcome. However, the ARD estimates for the no-schooling outcome suggest a narrowing of the gender gap, as the no-schooling rate for women in the 1986–1990 birth cohort is about 44 percentage points lower than that for women in the 1956–1960 birth cohort. The analogous estimate for men is –25 percentage points. Thus the ARD estimates indicate a narrowing of the gender gap.

The ARR and ARD estimates for some primary school, completion of primary school, and some secondary school are consistent with each other. That is, the ARRs and ARDs both point toward narrowing gender gaps. The outcomes measuring secondary school completion, completion of some college, and college completion indicate opposing patterns in the ARRs and ARDs: The ARRs for women are larger than they are for men, but the ARDs are larger for men than they are for women. The ARRs point toward narrowing gender gaps, while the ARDs suggest a widening of gender gaps. We reconcile the differences in the ARRs and ARDs by computing point estimates (marginal effects) for the gender gaps across the six schooling outcomes for each of the seven cohort groups. The computation of these point estimates allows us to present the evolution of the gender gaps for each educational outcome across all cohort groups. Thus one can easily detect trends (or a lack thereof). Ultimately, our analysis reveals the following conclusions: The gender gaps in rates of no-schooling, attending primary school, completing primary school, and attending secondary school have narrowed across cohort groups. However, for secondary school completion, some college, and college completion, we observe little to no progress.

2. Data and descriptive statistics

We use data from India's SES from 1983, 1987, 1993, 1999, 2004, and 2009 (Minnesota Population Center 2020) to measure progress or lack thereof in schooling outcomes for men and women. The long time span covered by the survey is useful (empirically) because the range of birth cohorts with overlaps in age increases with consistent surveying over long periods of time. We limit our sample to respondents between 22 and 26 years old. We then compute birth year using the survey year and the respondents' ages. We then group survey respondents into five-year birth cohort bins (1956–1960, 1961–1965, …, 1981–1985, and 1986–1990). Importantly, we observe survey respondents between 22 and 26 years old in each birth cohort.

Figure 1: Percentage in each education category by birth cohort group



Notes: The figure presents a stacked bar chart in which the individual education category indicates its share of the total. The percentages computed are based on the full sample, which includes 324,841 men and women who are 22–26 years old.

Respondents who did not report their education, gender, or age are excluded from the sample. The imposition of these restrictions results in a sample size of 324,841 observations. Although the numbers of observations associated with each birth cohort vary in size, the smallest sample size is 15,829 observations (the 1986–1990 cohort). The 1986–1990 cohort is smaller in large part due to only those aged 22–23 being observed in the cohort group. The remaining birth cohort groups consist of sample sizes that range

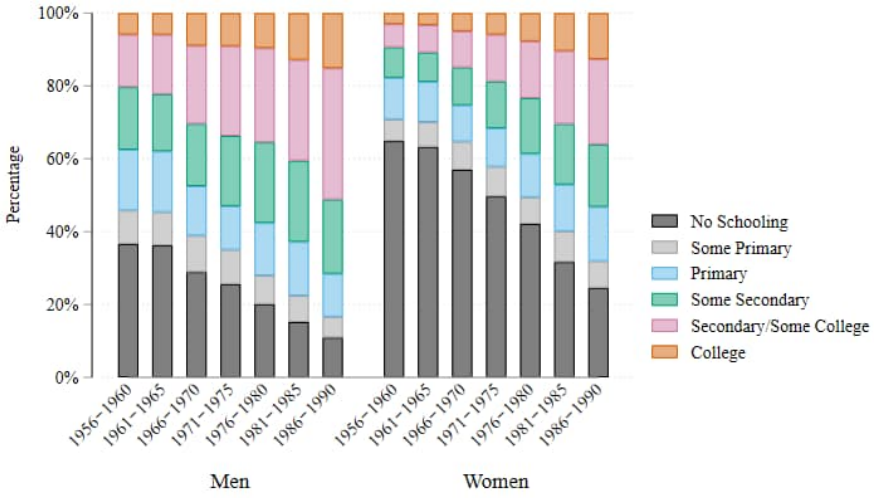
from about 40,000 to around 70,000. Thus there is ample statistical power to conduct empirical tests for cohort effects.

When pooling all cohort groups, we find that 31% had no schooling, 8% obtained some primary schooling, around 13% completed primary school, 17% had some secondary schooling, 21% completed secondary school/some college, and 9% completed a university degree. However, these percentages vary considerably across birth cohorts. In Figure 1, we report the relative frequencies of those aged 22–26 in each of the educational categories across the seven birth cohort groups. The figure scales the bars to be identical in height (100%). The sub-bars indicate the percentage contribution to the total for each education category. The share of respondents obtaining a secondary education or higher rises from about 18% for the 1956–1960 cohort group to around 51% for the 1986–1990 cohort group – a 33 percentage point increase. No-schooling rates plummet across the cohort groups as well, falling from around 45% for the 1956–1960 cohort group to around 12% for the 1986–1990 cohort group.

In Figure 2, we present the percentages of respondents who fall into each education category across birth cohort groups separately for men and women. Around 56% of men born to the 1986–1990 cohort group achieved at least a secondary education, while around 46% of women born to the same cohort group reached the same education milestone. For the earliest cohort group (1956–1960), approximately 12% and 23% of women and men, respectively, earned at least a secondary school degree. Between the latest and earliest cohort groups, the shares of men and women earning a secondary degree or higher increased by 33–34 percentage points. Despite similar percentage point differentials, the ascent of women relative to men in terms of secondary school completion diverges when it's evaluated as percentage changes (relative to a baseline outcome). If one expresses these computed differences as percentage changes, the magnitudes vary considerably: The rate of increase is 275% for women and 143% for men. Likewise, in terms of no-school rates, the gender gap is around 27 percentage points (in favor of men) for the earliest cohort group and 10 percentage points (in favor of men) for the latest cohort group.

Figures 1 and 2 suggest improvements in the gender gap in schooling outcomes in India. However, it is difficult in some cases, such as the secondary school completion/some college and college completion outcomes, to determine whether the gender gap across education outcomes has narrowed or widened. In response, we obtain point estimates for the gender gaps across all educational outcomes using a multinomial specification, which is described in the next section.

Figure 2: Percentage in each education group by birth cohort group and gender



Notes: The figure presents a stacked bar chart in which the individual education category is presented as a share of the total separately for men and women. The sample sizes for men and women are 161,892 and 162,949, respectively.

3. Empirical strategy

Our empirical tests are predicated on the idea that estimation of cohort effects, holding age effects constant, offers a direct test for progress or lack thereof. Identification of cohort effects requires that one avoid conflating age and cohort effects. Within a regression framework, one would need to account for both ages and cohorts to identify separately their effects on the schooling outcomes of interest. It is therefore necessary to observe respondents born to different cohorts at the same (or similar) ages, which is possible by pooling cross-sectional data collected in different years over a relatively long period (e.g., 20–30 years). The data used in our analysis are well suited for such an approach.

We use a multinomial specification to jointly model six education outcomes: (1) no schooling, (2) some primary school, (3) primary school completion, (4) some secondary school, (5) secondary school completion (includes some university or post-secondary training), and (6) university completion. The multinomial specification is useful in our context, as it is possible, perhaps likely, that the inter-cohort gender gaps vary across the

schooling outcomes considered.[3] The multinomial logistic specifications include three sets of explanatory variables: zero-one indicator variables for membership in the five-year cohort groups (the 1956–1960 cohort group is the omitted category), zero-one indicators for gender, and zero-one indicators for being 23, …, 25, and 26 years old (age 22 is the omitted category). In addition, the multinomial logit specification includes interaction terms between the cohort group indicator variables and the gender indicator variable. The estimates associated with these interaction terms indicate whether the gender gap is unchanged, widening, or narrowing across birth cohorts.

We are unable to account for year effects in the multinomial logit regressions, as the inclusion of controls for survey year results in perfect collinearity between the year and cohort group variables. Even if the data had been collected on an annual basis from 1983 to 2009, it would still not be possible to compute cohort effects between generations (approximately 30 years) while accounting for year effects, as the earlier cohorts would be observed only in the earlier survey years and the later cohorts in later survey years. For example, the 1956–1960 cohort would be observed from 1983 to 1989, and the 1986–1990 cohort would be observed from 2006 to 2009. We argue that the omission of these controls is largely inconsequential, as the "events" that take place between survey years do not bias our estimates but are instead potential explanations for the patterns in the data. Identifying underlying causes is outside the scope of this paper.

4. Results

In this section, we focus on empirically testing whether schooling attainment gaps narrow, widen, or remain unchanged across the cohort groups. We conduct two different sets of analyses using the same multinomial logit specification. First we estimate ARRs and ARDs for men and women separately (see Norton, Miller, and Kleinman 2013). Second we compute percentage point differentials between men and women within each cohort group and present confidence intervals around the estimates to gauge statistical significance.

[3] Likelihood ratio (LR) tests between the ordinal and multinomial logit specifications support modeling the educational outcomes as mutually exclusive outcomes in lieu of using an ordered specification.

Table 1: Adjusted risk ratios and adjusted risk differences for men and women – comparisons between the earliest (1956–1960) and latest (1986–1990) cohorts

	Men		Women	
	ARR	ARD	ARR	ARD
Education Outcome	(1)	(2)	(3)	(4)
No Schooling	0.3086	–0.2546	0.3886	–0.3978
	(0.0215)	(0.0089)	(0.0165)	(0.0115)
Some Primary	0.6349	–0.0334	1.2418	0.0143
	(0.0608)	(0.0059)	(0.1142)	(0.0066)
Primary	0.7145	–0.0477	1.3074	0.0350
	(0.0463)	(0.0081)	(0.0816)	(0.0090)
Some Secondary	1.1725	0.0296	2.0041	0.0840
	(0.0567)	(0.0095)	(0.1177)	(0.0089)
Secondary/ Some College	2.3270	0.1941	3.4144	0.1548
	(0.0919)	(0.0115)	(0.1673)	(0.0083)
College	3.0237	0.1120	4.8495	0.1096
	(0.1891)	(0.0089)	(0.3575)	(0.0077)

Notes: Based on a multinomial logit regression specification in which six education outcomes are regressed on indicator variables for gender, age, cohort groups, and interactions between gender and cohort group, we compute ARRs and ARDs using the adjrr command in STATA, developed by Norton, Miller, and Kleinman (2013). Conveniently, the package computes ARRs and ARDs for each outcome (we have six outcomes) in lieu of relative risk ratios, which compare to a base outcome. Furthermore, the package computes test statistics, such as standard errors (shown in parentheses) and 95% confidence intervals (omitted). The adjrr package, which builds on the margins package, allows one to compute ARRs and ARDs for men and women separately.

In Table 1 we report ARRs and ARDs, which are computed by comparing the earliest (1956–1960) and latest (1986–1990) cohorts for each of the six education outcomes.[4] ARRs less than 1 indicate that the outcome is less likely to occur between the latest and earliest birth cohort groups, whereas ARRs exceeding 1 indicate that the outcome is more likely to occur between the latest and earliest cohort groups. ARDs are percentage point differentials between the cohort groups, which are analogous to average treatment effects (Norton, Miller, and Kleinman 2013).

Using the ARRs from Table 1 (Columns 1 and 3), the percentage changes across the latest and earliest cohorts differ in magnitude between men and women: Compared to men born to the earliest cohort group, men born to the latest cohort group are about 69% less likely ($= 0.3086 - 1$) to not attend school; the same comparison for women yields an

[4] To implement these calculations, we use the statistical package developed by Norton, Miller, and Kleinman (2013). Their approach builds on previous work emphasizing the importance of computing risk ratios in lieu of odds ratios (e.g., Cummings 2009). Importantly, one of their extensions is the computation of ARRs and ARDs in the context of the multinomial functional form.

approximate 61% reduction (= 0.3886 – 1). In terms of primary school attendance and completion, there are notable gender differences. Among men, members of the latest cohort are 27% *less likely* to attend but not complete primary school. For women, the latest cohort group is about 24% *more likely* to attend but not complete primary school. The analogous estimates for completion of primary school are a 19% reduction among men and a 31% increase among women. For completion of some secondary school, the estimates imply a 17% increase among men and a 100% increase among women. For secondary school completion/some college, the estimates indicate a 132% increase among men and a 241% increase among women. Lastly, the percentage increases between the latest and earliest cohorts are large for college completion as well: 202% for men and 385% for women.

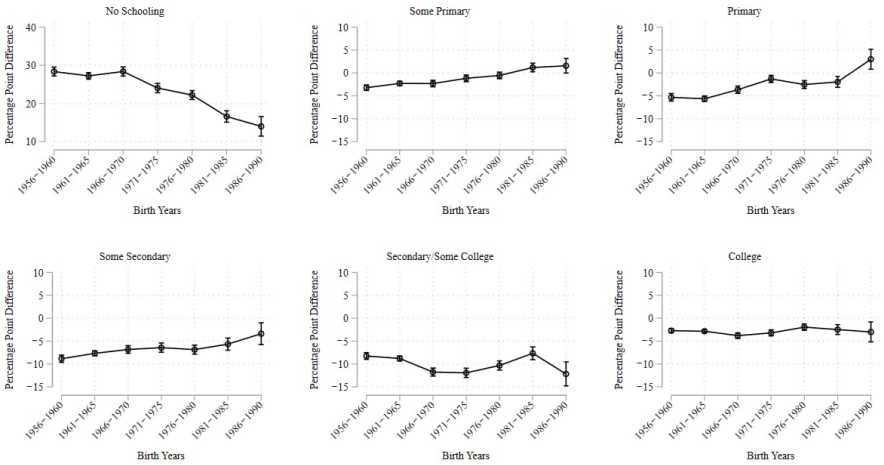
The ARDs give the percentage point differentials between the latest and earliest cohorts (Columns 2 and 4). The sign of the ARD is negative if the ARR is less than 1 and positive if the ARR exceeds 1. Rather than comment on each ARD estimate, we note the following: First, the ARDs for women tend to be larger (more negative for no-schooling rates and more positive for all other outcomes) than those for men. The two exceptions are secondary school completion/some college and college completion. For these outcomes, a comparison of ARRs for men and women suggests a narrowing of the gender gap across all education outcomes, except the no-schooling outcome. However, a comparison of the ARDs suggests a widening of the gender gap for the secondary school/some college and college completion outcomes as well as the no-schooling outcome.

In Figure 3, we reconcile these conflicting interpretations of the estimated ARRs and ARDs via estimation of marginal effects and plots of the point estimates (with confidence intervals) across the birth cohort groups. To compute the estimates, we rely on the same specification used to compute the ARRs and ARDs in Table 1. The *y*-axis of the six panels in Figure 1 is the percentage point differential. Negative (positive) values imply a gender gap in favor of men (women) over women (men).

For the no-schooling outcome, we observe a strong narrowing of the gender gap. For the earliest cohort, women are about 30 percentage points more likely than men to have no schooling. The estimated gender gap for the latest cohort is half the size of the analogous estimate for the earliest cohort (a 15 to 30 percentage point change in the estimated gender gap). The gender gap in completing some or all primary schooling reverses in sign between the earliest and latest cohorts in our sample. For example, the estimated gender gap for primary school completion is around –5 percentage points (men over women) for the earliest cohort but flips sign and is around +2 percentage points (women over men) for the latest cohort. The gender gap in the probability of completing some secondary schooling narrows across cohorts, but the gap is not eliminated. In particular, the gender gap for the earliest cohort is around –10 percentage points and the

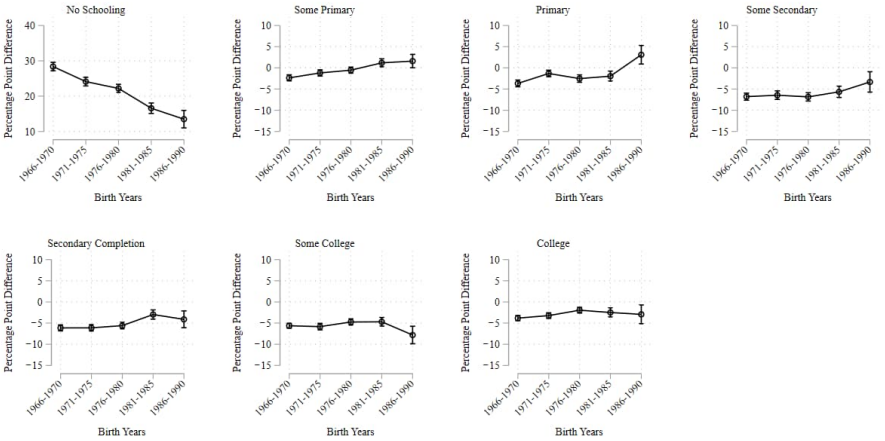
analogous estimate for the latest cohort is around –3 percentage points – a 7 percentage point reduction. The gender gap in secondary completion/some college widens between the latest and earliest cohorts: The initial gender gap is approximately –8 percentage points, and it rises to around –12 percentage points. We note, however, the sizable confidence interval around the estimate for the 1986–1990 cohort group for the secondary school completion/some college outcome. Despite the large confidence interval, the estimated difference in the gender gaps between the 1986–1990 and 1956–1960 cohort groups is statistically significant at the 5 percent level. Similarly, we observe a lack of gender progress in terms of college completion as well; that is, the estimated gender gaps for the different cohort groups are not different from each other in a statistical sense.

Figure 3: Gender gaps in education outcome across cohorts



Notes: The figure presents point estimates from a multinomial logit regression specification in which six education outcomes are regressed on indicator variables for gender, age, cohort groups, and interactions between gender and cohort groups. The sample size consists of 324,788 observations. The *y*-axis measures the percentage point difference in the education outcome between men and women for each cohort. Confidence intervals (95%) are presented around the estimates. For the no-schooling outcome, a downward (upward) sloping function implies a narrowing (widening) of the gender gap, while an upward (downward) sloping function implies a narrowing (widening) of the gender gap.

Figure 4: Gender gaps in education outcome across birth cohort groups robustness check with subsample based on data from 1993–2009



Notes: The figure presents point estimates from a multinomial logit regression specification in which six education outcomes are regressed on indicator variables for gender, age, cohort groups, and interactions between gender and the cohort groups. The sample size consists of 204,529 observations. The *y*-axis measures the percentage point difference in the education outcome between men and women for each cohort. Confidence intervals (95%) are presented around the estimates. For the no-schooling outcome, a downward (upward) sloping function implies a narrowing (widening) of the gender gap, while an upward (downward) sloping function implies a narrowing (widening) of the gender gap. The only difference between Figures 3 and 4 is that in Figure 4, data collected prior to 1993 are excluded.

A data limitation facing our analysis is the lumping together of secondary completion and some college into a single category. In particular, the 1983 and 1987 SESs do not distinguish between secondary school completion and some college. However, in the 1993 and subsequent surveys, the SESs refined the education question to include separate entries for secondary school completion and college attendance. As a sensitivity check, we exclude observations from 1983 and 1987 and re-estimate the multinomial specification of interest. The consequence of losing the 1983 and 1987 samples is a reduction in statistical power and fewer cohort groups eligible for inclusion in the analysis. For this subsample, the 1966–1970, …, 1981–1985, and 1986–1990 cohorts are included; thus the 1956–1960 and 1961–1965 cohort groups are excluded from the analysis.

For the subsample, the estimated gender gaps for each of the seven education outcomes across the different cohort groups are presented in Figure 4. Splitting the secondary completion/some college variable into separate measures does not alter the conclusions from Figure 3 in a material way. We observe a narrowing of the estimated gender gaps across cohort groups for all education outcomes below secondary completion. For the secondary completion outcome, we observe a generally upward-

sloping pattern across cohort groups, but statistical imprecision complicates inferring a narrowing (or widening) of the gender gap. For the some college outcome, we detect a general widening of the gender gap across cohort groups, although the gender gap for the 1986–1990 cohort group is estimated imprecisely. The evolution of the estimated gender gaps across cohort groups does not materially change for the college completion outcome.

5. Conclusions

Despite substantial improvements in literacy rates for India in the last several decades and a narrowing of the gender gap in terms of years of completed schooling, we still lack a comprehensive understanding of how gender disparities have evolved for different education milestones (e.g., graduating from secondary school instead of attending but not finishing). Relying on a cohort study framework that studies young people (22–26 years old) and a multinomial logit specification, we investigate whether the gender gap in the education outcomes has narrowed, widened, or remained unchanged across the 1956–1960, 1961–1965, …, 1981–1985, and 1986–1990 birth cohort groups. Our findings are mixed in the sense that we detect a narrowing of the gender gap when considering education outcomes on the lower end of the schooling spectrum (some primary school, completing primary school, and some secondary school), whereas we demonstrate little to no change in the gender gap across education outcomes toward the upper end of the education spectrum (completing secondary school, attending but not completing college, and graduating from college).

Although we observe improvements in the gender gap in schooling outcomes toward the lower end of the education spectrum, gender inequities associated with higher levels of schooling are persistent across cohort groups. It is important to understand the causes of these patterns, as there are likely important policy considerations for India as it grapples with the interactions among technological change, a relatively young workforce, and persistent gendered norms and attitudes.

References

Cummings, P. (2009). The relative merits of risk ratios and odds ratios. *Archives of Pediatrics and Adolescent Medicine* 163(5): 438–445. doi:10.1001/archpediatrics.2009.31.

Dhar, D., Jain, T., and Jayachandran, S. (2018). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. (National Bureau of Economic Research Working Paper Series 25331). Cambridge, MA: NBER. doi:10.3386/w25331.

Duraisamy, P. (2002). Changes in the returns to education in India, 1983–94: By gender, age-cohort, and location. *Economics of Education Review* 21(6): 609–622. doi:10.1016/S0272-7757(01)00047-4.

Katiyar, S.P. (2016). Gender disparity in literacy in India. *Social Change* 46(1): 46–69. doi:10.1177/0049085715618558.

Kingdon, G.G. (2007). The progress of school education in India. *Oxford Review of Economic Policy* 23(2): 168–195. doi:10.1093/oxrep/grm015.

Minnesota Population Center (2019). Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN: IPUMS. https://international.ipums.org/international/.

Mitra, A. (2019). Returns to education in India: Capturing the heterogeneity. *Asian and Pacific Policy Studies* 6: 151–169. doi:10.1002/app5.271.

Norton, E.C., Miller, M.M., and Kleinman, L.C. (2013). Computing adjusted risk ratios and risk differences in Stata. *The Stata Journal* 13(3): 492–509. doi:10.1177/1536867X1301300304.

Self, S. and Grabowoski, R. (2004). Does education at all levels cause growth? India: a case study. *Economics of Education Review* 23(1): 47–55. doi:10.1016/S0272-7757(03)00045-1.

World Economic Forum (2020). Global Gender Gap Report 2020. http://www3.weforum.org/docs/WEF_GGGR_2020.pdf.