The Stubborn Law of Female-Male Literacy: Why the Gap May Widen

Maneesh Arora and Rein Taagepera
University of California, Irvine
3151 Social Science Plaza
Irvine, CA 92697, USA.

Abstract

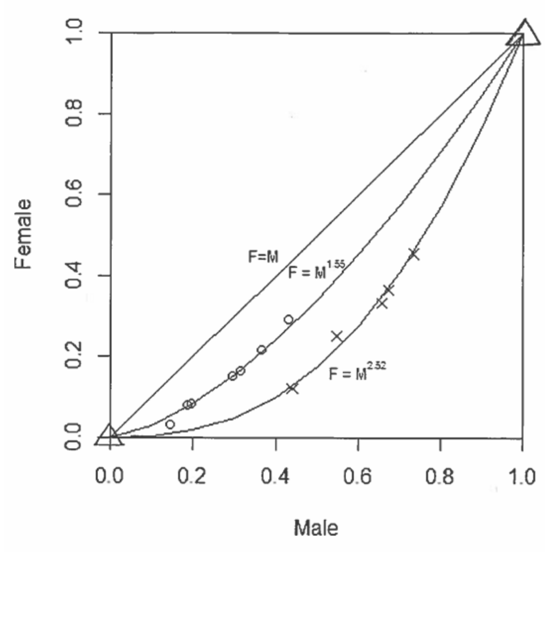
Graphing female literacy rates (F) versus the male (M) reveals an empirical law of female-male literacy: "As literacy in a country increases, female literacy rate remains a fixed exponent of the male" – $F=M^k$. Hence the male-female literacy gap is bound to widen at first and shrink later on. This may lead to a rethinking of literacy policies, particularly in those countries with low overall literacy rates. Values of exponent k range from 1.5 to 2.5. They may express deep-set gender bias that persists throughout modernization, even while UNDP Gender Inequality Index may decrease. The broadest implication is that locating deterministically couched laws, analogous to those in natural sciences, needs conceptual thinking rather than only statistical analysis

Keywords: Female literacy; male-female literacy gap; UNDP Gender Inequality Index; deterministically couched laws.

Introduction

It has been common knowledge that female literacy rate trails the male, at least until high levels are reached. But we should not be satisfied with knowledge about mere existence and direction of this gap. This would be akin to being satisfied with knowing that things fall downwards, and feeling no need for equations of motion devised by Galileo. Yet the relationship between female and male literacy rates seems not to have been investigated comparatively in any depth. It may have been assumed that this relationship is haphazard and not subject to any universal regularity. However, we find evidence that it is. As literacy in a given country increases over time, we find that a pattern close to $F=M^k$ is maintained, where k is a constant specific to the country. (Both female literacy rate F and male literacy rate M must be in fractional shares out of 1, not percent.) In the case of 15 countries that started with F less than 0.23 around 1980, the exponent k ranges from 1.5 to 2.5. Figure 1 shows sample curves F vs. M at low and high ends of k: Burkina Faso (k=1.55) and Mozambique (k=2.52).

Figure 1: Female vs. male literacy in Burkina Faso 1975-2015 and Mozambique 1980-2015, and curves $F=M^1$, $F=M^{1.55}$, and $F=M^{2.52}$. Also shown are conceptual anchor points 0,0 and 1,1.



While some countries show less clear-cut patterns than these two countries, the overall picture will be shown to be sufficiently uniform to suggest the existence of a social law, the law of female-male literacy: *As literacy in a country increases over time, female literacy rate grows as a fixed exponent of male literacy rate – $F=M^k$.*

This law is an empirical one, at this stage of inquiry. It is begging for logical justification in terms of a specific process of interaction between male and female literacy. The present study, however, limits itself to establishing the empirical existence of this relationship.

An important policy consequence results. Reducing the male-female literacy gap has received considerable attention. Countries have been concerned about this gap persisting or even widening. They are relieved when the gap finally begins to narrow down, as it has happened in India since 1981 (Infochangeindia, 2015a). It might seem that efforts to boost female literacy rate are finally paying off. A different reason, however, shows up in Figure 1, where this gap is the vertical distance between the line $F=M$ and the curve $F=M^k$. Indeed, if the pattern $F=M^k$ applies, then the male-female literacy gap is bound to widen at first and shrink later on. For Burkina Faso the gap has tended to widen not because of official neglect but because the country has been in the early stage of developing literacy, where $F=M^k$ implies a widening gap. In contrast, Mozambique has by now reached the late stage of developing literacy, where $F=M^k$ implies a shrinking gap, regardless of policy emphasis. Stubbornly**,** the law $F=M^k$ seems to impose itself in most countries, with a constant value of k deeply embedded in their culture. The up-and-down of the male-female literacy gap then mechanically follows. We'll return to this important and possibly controversial claim.

First, we survey past work on female/male literacy rate. Second, we present the theoretical framework – the logical constraints that data fits must respect. Third, we discuss the data. Fourth, we analyze these data, with emphasis on agreement or disagreement with format $F=M^k$. Fifth, we discuss the impact on the male-female literacy gap. Sixth, we look for connection with the general UNDP Gender Inequality Index, but instead, observe a connection to the main European language used. Finally, we discuss the broader implications of this study and the method it uses.

2. Literature Review

While it has been common knowledge that female literacy trails the male (at least until high levels are reached), little analysis has been carried out. Indeed, even data on female and male literacy rates separately are hard to come by for times prior to 1970. Our review of the literature on literacy rates and literacy development turned out little data and even less quantitative analysis. Qualitatively, many scholars discuss the theoretical development of literacy rates, and discuss the differences in female and male literacy development. However, we found no evidence of work done to show quantitatively how female and male literacy develop together. Some scholars have graphed female and male literacy rates over time (but usually not versus each other) for particular nations or sub-regions.

Thus Sharma (1987) graphs the male, female and overall literacy rates in India (excluding Assam) against time, from 1901 to 1981 using data from the Office of the Registrar General & Census Commissioner. He fits these three data sets with three straight lines (the extrapolation of which would lead to sub-zero literacy rates in the far past and over 100% in the presumable far future). Sharma also calculates the ratio of female to male literacy rate over time and graphs several socio-economic and demographic variables by female literacy rate but does not attempt to fit any of the data patterns with a curve. He does not graph female vs. male literacy rate. We have found only two cases of graphing female vs. male literacy rate. Dreze and Loh (1995) graph F vs. M for Indian states and Chinese provinces in 1981-82 and 1991, providing a useful visual of educational disparity. They draw the conclusions that female literacy is significantly lower than male literacy in both countries and that the disparity is particularly striking in India. They establish that there is some regularity but do not attempt to fit the data pattern with a curve, much less to establish a corresponding equation.

Huebler (2007) graphs F vs. M for 135 countries, based on contemporary UNESCO (June 2007) data. He shows the line $F=M$ as a comparison basis, but he does not try to fit the data pattern with a curve. There are good qualitative explanations for the relationship between female and male literacy development, including social and gender biases and national development. Some of the following studies deal explicitly with literacy, while others deal with education more broadly, with implications for literacy.

2.1 Social Biases

Lori McDougall (2000) finds a strong correlation between regions in Uttar Pradesh that have narrowing gender gaps and those with rising female literacy rates. McDougall shows that many of the benefits associated with greater female literacy are societal, including a more productive workforce, lower fertility, and lower infant mortality. The costs, on the other hand, are often private. One of the main costs associated is the foregone opportunity cost of child labor. In India, like many other countries, parents may view the benefits of keeping their female children at home to work in the house or in the field as outweighing the benefits of that child attending school. That view is less salient in regards to male children.

McDougall finds evidence of other social biases leading to gender gaps in literacy development, including the following: social restrictions on female workforce mobility; fear that educated females will be less attentive to household chores and less likely to obey their parents and husbands; cost of education; and teacher bias against female students. Together, these social and gender biases have kept female enrollment in schools and literacy rates lower than that of males. As biases slowly dissipate female education and literacy increase faster and the gender gap starts to shrink. Raju (1988), in another study of literacy rates in India, has similar findings. He also attributes the rise in female literacy rates to decreases in gender bias. He argues that the spread of Christianity and the resulting spread of Christian values throughout India have advanced egalitarianism and modernization, thus decreasing gender biases and increasing female literacy rates.

2.2 National Development

King and Hill (1993) find strong connections between female education and national development. The level of female education tends to be lowest and the gender gap tends to be highest in the poorest countries. Female enrollment in primary and secondary schools has increased much faster between 1960 and 1988 in upper-middle income countries than lower-middle and lower income countries. A gender gap in enrollment rates exists in all three categories, but has diminished the most in upper-middle income countries. Hill and King also find large regional differences in enrollment rates in developing nations. East Asia and Latin America have achieved close to equal enrollment in primary schools for females and males, while the other regions lag behind. In South Asia, the gender gap in primary school enrollment has actually widened. The overall trend, however, is for enrollment rates to become more equal between females and males. National development and resulting modernization enhance education enrollment and literacy rates of both females and males. Rising incomes and higher levels of development throughout the world have led to rising literacy rates.

2.3 Female Achievement

Although females have historically been disadvantaged in educational and literacy opportunities, and continue to be disadvantaged in much of the world, studies have shown that when given more equal opportunities females often outperform males in reading ability. Klein (1977) finds that, when using standardized reading tests and achievement tests as measures, female scores have traditionally exceeded male scores. Similarly the Saskatchewan Department of Education assessed the reading and writing abilities of 5th, 8th, and 11th grade students living in rural areas. Their findings show that female students at all three grade-levels were significantly stronger in reading ability (Saskatchewan Education, Training and Employment, 1996). A broader study by the International Association for the Evaluation of Educational Achievement (IEA) found that among the 32 countries that were part of the study, reading scores tended to favor females (Gamble and Hunter, 1999). When given equal opportunities, females not only equal male literacy abilities, but also often surpass them. In sum, our survey of the literature makes us conclude that while several qualitative explanations of female and male literacy development have been suggested, the quantitative relationship between female and male literacy rates seems not to have been graphed and analyzed comparatively in any depth.

3. Theoretical Framework

Given data like those for Burkina Faso and Mozambique in Figure 1, how should we proceed to analyze them? One could "simply" fit the data with straight lines ($F=a+bM$), using OLS or other statistical approaches to linear regression. But this would be a dead end. These lines would be approximately $F=-0.09+0.90M$ for Burkina Faso and $F=-0.35+1.10M$ for Mozambique (not shown in Figure 1), with superb correlation – R^2 around 0.9. However**,** when adding further countries, we would get a maze of values of constants a and b, with no systematic relationships between them.

While good for interpolations within the range of data points, outside this range these regression equations would be catastrophically off. Indeed, they would claim that, when male literacy was near 0 in Burkina Faso, its female literacy rate would have been -0.09 (*minus* 9%). For Mozambique, the female literacy rate would have been -0.35 (*minus* 35%). At the other end, even with male literacy at 100%, female literacy rate could not reach beyond 0.81 (81%) in Burkina Faso and 75% in Mozambique. The reality is that, with near zero male literacy, all countries have near zero female literacy. And with near 100% male literacy, all countries have near 100% female literacy. If we want to consider their relationship throughout their possible range, 0 to 1 (100%), linear statistical fits will not do. If we seem here to belabor the obvious, this is because social scientists have (mis)used linear regression all too often in similar situations. We obtain more insights by logical considerations, before applying statistical data fits. Logic must inform the format of statistical analysis, if the latter is to make sense.

The present quest is guided by broad logical considerations. In the distant past when male literacy in a country was zero, it is hard to assign female literacy any other number than also zero, given that we observe that female literacy almost never surpasses the male. More formally: If $F \leq M$, then $M=0 \rightarrow F=0$. This would be the point 0,0 at the bottom left corner in Figure 1. This represents a *conceptual anchor point* (Taagepera, 2008: pp. 34-37; and 2011), valid in a distant past. Here $M=F$. Conversely, when even female literacy approaches 1, it is hard to assign male literacy any other number than also approaching 1: If $M \geq F$, then $F=1 \rightarrow M=1$. This would be the point 1,1 at the top right corner in Figure 1. This is another conceptual anchor point, toward which countries are moving as literacy increases. Here again $M=F$.

As literacy increases over time from 0 to near 1, the point F,M must trace a continuous path from one conceptual anchor point to the other – from 0,0 to 1,1. The straight line marked $F=M$ would be such a pattern, if female and male literacy moved up apace, over time. When female literacy lags the male, the pattern must be curved. Which path is literacy likely to take? As Albert Einstein is reputed to say: Our models must be as simple as possible – but not simpler. In the present case, $F=M^k$ is the simplest way to travel from 0,0 to 1,1, subject to $F \leq M$. This is the simplest in two senses. First, $F=M^k$ is the shortest expression possible, and second, it involves the least number of adjustable parameters – only one, k. (In contrast, note that the "simple" linear data fit $F=a+bM$ is more complex, as it involves two parameters, a and b.) So this is the format to be tried first. We should look for more complex formats only if data contravened $F=M^k$ – or if further logical considerations demanded so. Does $F=M^k$ supply the best possible statistical fit for actual data? Most likely this is not so. One might do somewhat better with a two-parameter equation $F=aM^k$, where a could take values different from 1. But this would be statistically wise, logically foolish. Indeed, $a<1$ would imply that F could never reach 100%, not even at $M=1$. And $a>1$ would imply that, at $M=1$, F would reach beyond 100%. Our models and data fits must not predict absurdities, if we want to be taken seriously as scientists.

4. Data

Data for female and male literacy rates separately are quite sparse prior to the late 1970s. Which countries should we use to test the validity of $F=M^k$? We must consider the range of random error. If a country perfectly fitted a curve $F=M^k$, then k would be the same for all data points: $k=\log F/\log M$. But random error can distort this value, and its range is especially wide when M and F approach 0 or 1. Therefore, very low and very high values of F (say, $F<.03$ or $F>.90$) should be avoided. The lower limit .03 excludes no countries since mid-1970s, while most countries by now do reach the higher limit .90. We took as our data set those countries in UNESCO Institute for Statistics (2015) where the earliest recorded female literacy rate was less than 23% ($F<0.23$), and data were available for at least 3 different years spaced far apart. These 15 countries, all Afro-Asian, are shown in Table A in the Appendix. In addition to M and F, this table also shows the ratio $k=\log F/\log M$ for each data point.

This data set is far from ideal. Countries may have different definitions of literacy, which can cause literacy rates to be unreliable. There is no objective literacy test used internationally, which can cause reporting biases and makes literacy rates difficult to compare cross-nationally. However, all of the literacy rate data used in our study comes from the UNESCO Institute for Statistics (2015) which attempts to standardize literacy rates to adults 15 years and older. Additionally, literacy rates are consistent over time, which causes one to suspect that the rates are generally reliable (Dreze and Loh, 1995). Some countries visibly have upgraded their criteria, most markedly Central African Republic, where both M and F drop steeply from 2000 to 2010 (M from 0.668 to 0.507, and F from 0.352 to 0.244). Note, however, that this drop affects both M and F, so that the resulting $k=\log F/\log M$ remains in the usual range (though decreasing from 2.6 to 2.1). This is the case more broadly:

Criteria may change, and data may be poorly collected and mildly doctored, but the value of k=log$F$/log$M$ remains rather resilient. Another problem with the data is that figures are rather sparse in the 1970s and 1980s, while in the 2000s they sometimes are available for consecutive years (at times with diverging k values), followed by longer gaps. We accept all data available, rather than try to cull the data and risk selecting in favor of our favorite format.

As it is, the data range from 1975 to 2015 for Burkina Faso, while only from 1988 to 2015 for Senegal. The values of M range from 0.135 (Mali 1976) to 0.888 (Burundi 2008). The values of F range from 0.032 (Burkina Faso 1975) to 0.846 (Burundi 2008). Thus the data cover most of the entire range from 0 to 1, excluding the extremes, where k is subject to large random error.

5. Analysis

On the basis of data in Table A we calculated three indicators for every country, as shown in Table 1: the overall mean of exponent k in $F=M^k$, and separately, the means for the first and second halves of data over time, respectively designated as k_- and k_+. (For odd number of data points, the central one is averaged into both, at half weight.) The reason is the following. If a country truly follows $F=M^k$, then k_- and k_+ are equal ($k_-=k=k_+$). An actual difference k_--k_+ (positive or negative) would reflect the degree of deviation from $F=M^k$.

Table 1: Partial and overall means of exponent k in $F=M^k$, UNDP Gender Inequality Index, and main European language. Countries are ranked by increasing mean k.

Country	k_-	k	k_+	GII	Main European Language
Bangladesh	1.70	1.49	1.29	0.529	English
Burkina Faso	1.59	1.55	1.52	0.607	*French*
Mali	1.57	1.65	1.74	0.673	*French*
Burundi	1.83	1.66	1.48	0.501	*French*
Benin	1.86	1.88	1.89	0.614	*French*
Senegal	1.76	1.93	2.09	0.537	*French*
Morocco	2.15	2.16	2.16	0.460	*French*
Centr. Afr. R.	2.19	2.19	2.18	0.654	*French*
Egypt	2.34	2.19	2.04	0.580	English
Pakistan	2.19	2.25	2.30	0.563	English
Guinea-Bissau	2.29	2.26	2.21	----	Portuguese
Nepal	2.27	2.28	2.28	0.479	English
Afghanistan	2.41	2.29	2.18	0.705	English
Liberia	2.13	2.34	2.54	0.655	English
Mozambique	2.47	2.52	2.56	0.657	Portuguese

Calculated from data in Table A. UNDP Gender Inequality Index from United Nations Development Programme Human Development Reports (2014).

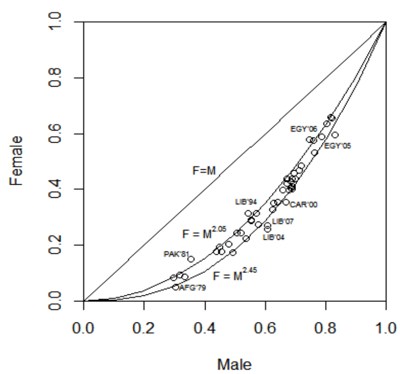
This k_--k_+ turns out to be balanced (positive in 7 cases and negative in 8) and limited: In 7 cases the deviation from 0 is within ±0.10. Only 4 countries deviate more than that in the positive direction and another 4 in the negative. The mean of k_--k_+ from Table 1 is extremely close to 0: +0.02. This means that, on the average, the model $F=M^k$ holds up remarkably well. The major deviant cases (absolute values of k_--k_+ larger than 0.30) balance themselves out: on the positive side Bangladesh (+.41), Burundi (+.35) and Egypt (+.30); on the negative side Liberia (-.41) and Senegal (-.33). We'll consider each of them in some detail.

Bangladesh deviates systematically from model $F=M^k$: Over time, its k values steadily keep falling. Ever since 1991 its female literacy rate has kept increasing sharply, while its male literacy rate has stalled, comparatively. If the trend of the past 20 years continued, female literacy in Bangladesh would catch up with the male in the late 2020s, at $F=M$=0.70, and then surpass it. This is foreshadowed in youth (15-24 years) literacy rates, where females surpassed males 80.4 to 77.1% in 2008-2012 (UNICEF, 2015). Such extrapolation is risky, but it highlights the highly unusual pattern of Bangladesh. *Burundi* follows a similar pattern, but with even female literacy already around 84% future projections are less spectacular.

In contrast, *Liberia* has its *k* peaking in the middle of the time period and then it decreases again, suggesting large but random variation in the way literacy is reported. From 1994 to 2004, male literacy keeps increasing from 55.4% to 60.8%, yet female literacy inexplicably drops from 31.3% to 25.7%! Hence Liberia does not deviate systematically from $F=M^k$. *Egypt* follows a similar pattern in the reverse direction. Its *k* peaks in 2005 when reported male literacy jumps to a level (83.0%) never reached thereafter, while female literacy keeps increasing rather steadily. In *Senegal* random variation is appreciable. On the whole, the gap between male and female literacy has kept increasing at a male literacy level (in the 60%) where the gap usually begins to close. Accordingly, *k* has tended to increase.

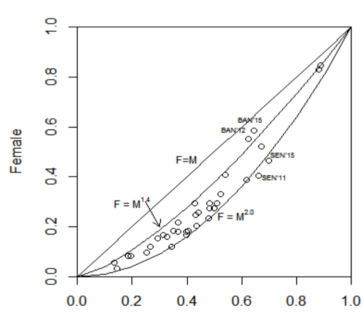
The mean values of *k* happen to bunch into two groups, plus an outlier, shown separated in Table 1. There are 6 countries with *k* ranging from 1.5 to 1.9, 8 countries with *k* ranging from 2.15 to 2.35, and Mozambique at k=2.52. Within these groups countries have essentially the same *k*: k=1.7±0.2 and k=2.25±0.10. So we'll graph them as if they represented random variations around the same culturally determined value of *k*. We start with the closest-knit group. The country means of *k* range from 2.15 to 2.35. Allowing for random fluctuation of just ±0.10 around these means widens the expected range of individual data points to $2.05<k<2.45$. Figure 2 shows curves delimiting this zone, plus data points from Table A. Most points locate indeed within the expected zone, filling it rather uniformly. Major deviations above the upper curve (meaning extra low *k*) are labeled in Figure 2: Pakistan 1981, Liberia 1994 and Egypt 2006. Below the lower curve (meaning extra high *k*) are Afghanistan 1979, Liberia 2004 and 2007, CAR 2000 and Egypt 2005. We observed earlier that Egypt and Liberia show large random variation in both directions.

Figure 2: Female vs. male literacy for 8 countries where mean *k* ranges from 2.15 to 2.35, and curves $F=M^{2.05}$ and $F=M^{2.45}$.



For the other group the country means of *k* range more widely, from 1.5 to 1.9. Allowing for random fluctuation of just ±0.1 around these means widens the expected range of individual data points to $1.4<k<2.0$. Figure 3 shows curves delimiting this zone, plus data points from Table A. Most points again locate within the expected zone, filling it rather uniformly. Major deviations above the upper curve are Bangladesh 2012 and 2015. Major deviations below the lower curve are Senegal 2011 and 2015. We observed earlier that these countries were problematic.

Figure 3: Female vs. male literacy for 6 countries where mean *k* ranges from 1.5 to 1.9, and curves $F=M^{1.4}$ and $F=M^{2.0}$.



Outside these groups, Mozambique consistently has very high *k*. It fits the model $F=M^k$ well (cf. Figure 1).

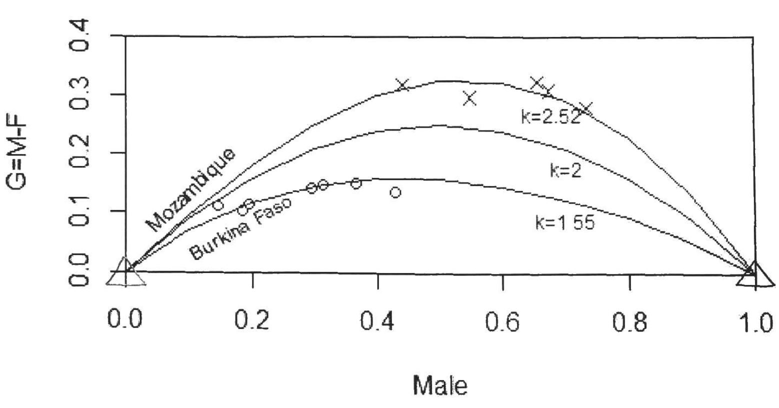
In sum, systematic deviations from a steady exponent k are few and balance themselves out. Only Bangladesh and possibly Burundi have clearly decreasing k. Deviations are large but random for Liberia, Egypt and Senegal. Thus, on the whole, the model $F=M^k$ holds. Consequently, we claim the existence of an empirical law of female-male literacy: *As literacy in a country increases over time, female literacy grows as a fixed exponent of male literacy – $F=M^k$*. Rarely do other factors override the workings of this law in a marked way. Is $F=M^k$, with k ranging from 1.5 to 2.5, a peculiarity of present times, or is it valid more broadly? The only good data series from earlier times seems to be India, starting with 1951 (M=.272, F=.089, k=1.86) (Infochangeindia, 2015b). Apart from this first date, India's data fall comfortably into the range delineated in Figure 2.

For much earlier times, scholars offer samples of data on ability to sign one's name in North America and Europe. In America's New France 1663, this was .59 and .46 for grooms and brides (k=1.47). For men and women, it was .42 and .30 (k=1.39) in 1680-1699, and a surprising .45 and .43 (k=1.06) in 1657-1715 (Magnuson, 1985). In England 1841, it was .67 and .56 (k=1.45). The criteria are fuzzy, and we have no time series for the same population, so stability of k over time cannot be verified. The values of k are below those observed in recent times in Africa and Asia, suggesting a smaller male-female gap at a given level of male literacy. This may be expected in those Protestant countries where women as well as men were urged to read the holy script, but it is more surprising for the mostly Catholic French.

5. Male-female gap in literacy – and Education

What does the law of female-male literacy imply for policies and campaigns to advance female literacy, so as to reduce the male-female literacy gap? To address this question, we alter Figure 1 so as to bring this gap into evidence. Figure 4 shows the gap $G=M-F$ graphed against M, both for Burkina Faso and Mozambique, along with the curves $G=M-M^{1.55}$ and $G=M-M^{2.52}$, which are expected to fit the data. Consider the curve that fits Burkina Faso data. The gap is zero at M=0 and then widens, as M increases faster than F. The gap reaches a maximum around M=0.5 and then begins to decrease. As near-complete literacy is reached (M,F>0.90) the gap becomes tiny. (Indeed, it sometimes reverses itself, as slightly more men than women may still remain illiterate.) The same basic gender bias, reflected in the value of k, would lead to different literacy gaps at different levels of technological development. Extrapolating from the graph, it can be expected that in Burkina Faso the gap will remain at the present level of about 15 percentage points for a long time, until M reaches 60% (and F reaches 45%), no matter what one does to boost M and F separately. Only thereafter will the gap begin to shrink. Mozambique, on the other hand, has reached this stage, and its male-female gap will decrease steeply with increasing M. (How fast M will increase is not predicted.)

Figure 4: Male-female literacy gap vs. male literacy in Burkina Faso 1975-2015 and Mozambique 1980-2015, and the curves corresponding to $F=M^1$, $F=M^{1.55}$, and $F=M^{2.52}$. Also shown are conceptual anchor points 0,0 and 1,0



Policies and campaigns to advance female literacy can speed up the move toward higher female literacy, but the law $F=M^k$ indicates that in so doing female literacy campaigns also impact male literacy. Indeed, when male literacy is below 40%, they could boost male literacy even more than the female. How could such a paradoxical outcome come about? With so many illiterate males still around the targeted women, those men may be activated even more than the women themselves. We do not find this scenario very convincing, but otherwise $F=M^k$ could not be maintained the way it is (outside Bangladesh and possibly Burundi and Egypt).

Conversely, when male literacy reaches 60%, female literacy will begin to catch up with the male even in the absence of concerted effort. If an effort to reduce the gap is made, it may appear as though this effort "finally begins to pay off", when actually it's a matter of maintaining $F=M^k$. At this stage, even a conceivable campaign to boost male literacy would actually reduce the male-female gap: With few illiterate males left, women may be activated more than the men themselves.

Decreases in male-female literacy gap have often been attributed to decreases in gender bias (Raju, 1988; McDougall, 2000). But the constancy of exponent k suggests another possibility: a rather stable basic gender bias reluctantly interacting with a desire to modernize. As more men become literate, literacy loses its elite character, which facilitates greater access to literacy for women despite continuing bias. In sum, to reduce the male-female literacy gap, general policies and campaigns to boost literacy may be as effective as those targeting women specifically. It's just a matter of getting male literacy rate above 60% or so, and the female rate will start rising "miraculously".

The case of Bangladesh shows that other factors can counteract the law $F=M^k$, even while it rarely happens. Usually, a constant value of exponent k seems deeply embedded in a culture. It remains to be elucidated which factors in Bangladesh (and possibly Burundi and Egypt) have reduced the value of k over time. Female literacy rate slightly surpasses the male in quite a few countries when both reach very high levels: The gap $G=M-F$ becomes slightly negative. But as Huebler (2007) points out, two countries have a highly negative G, and at a rather low M: Lesotho (M=.737, F=.903, hence G=-.166 and k=0.334 – much below 1) and Jamaica (M=.741, F=.859, G=-.118 and k=0.507). Bangladesh may be headed in such a direction.

Previous observations about increasing and decreasing literacy gaps in various countries and regions (Raju, 1988; King and Hill, 1993; McDougall, 2000) now fall into a common pattern, rather than being confusing and contradictory. But these studies also open wider vistas. King and Hill (1993) have pointed out that in South Asia the gender gap in primary school enrollment has widened, even while total enrollment has increased. On the other hand, the gender gap in enrollment rates at all levels of education has diminished the most in upper-middle income countries. This reminds us that the format $F=M^k$ may apply to gender disparities other than literacy. Addressing this would be a separate research project.

6. Exponent k as a measure of deep-set persistent gender bias

As advancing technology brings about socioeconomic changes, the prestige value of various skills is altered. Being literate is a distinction when few others are. It stops being so when practically everyone else also is literate. At this stage, there is no longer any point in denying literacy to social categories against which dominant views harbors a bias. This does not mean that bias has gone away but only that it retrenches itself at a different point of application.

In a society with deep-set gender bias, female literacy may be resisted until male literacy becomes so prevalent that few males feel threatened by literate women. Yet at the same time they might still object to higher education for women. Thus there could be Protestant countries that insisted on female literacy so that they could read the Holy Book, yet excluded women from universities. Once the economic demands of a high-tech society made this barrier pointless, the visible or glass ceiling would move elsewhere. One may wonder whether the exponent k is a measure of this deep-set gender bias in a society. This is an underlying bias that remains rather constant even while literacy rates rise (and by this very rise devalue the relative prestige of being literate).

In contrast, some other measures of gender inequality make use of criteria significant at the present stage in world history. The present female literacy rate or male-female literacy gap may be examples – or indeed, any measure based on the present levels of indicators, such as the UNDP Gender Inequality Index (GII). They are important and useful – as measures of the present situation. Exponent k measures something else that transcends time. It shouldn't then come as a surprise (although it initially came to us!) that the current values of GII (also shown in Table 1) are utterly unrelated to k. The countries in the low-k group have no appreciably lower GII (average .58), compared to the higher-k group plus Mozambique (average .59). Conversely, the average k for 7 lower-GII countries (1.99) is hardly lower than for 7 lower-GII countries (2.06). Overlooking Bangladesh with its unusual pattern, the extremes of k are Burkina Faso and Mozambique. In what do they differ? For one, it's French vs. Portuguese colonial past and the resulting European language being used most. Table 1 indicates the present main European language used, largely due to past colonial domination or influence.

Apart from Mozambique, k tends to be highest for countries that escaped colonial conquest (Liberia, Afghanistan, Nepal), and it is lowest for former French colonies. The pattern is even clearer regarding language: For k<2.19 it's French (except for Bangladesh), while for k>2.19 it's English and Portuguese. Why would the main European language affect deep-set gender bias in Afro-Asian countries? Note first that the main European language is another deep-set historical given that changes only slowly (e.g., as English penetrates formerly Francophone areas). It may exert a permanent influence on ways of thinking and on current sources of outside information. In many ways Anglophone and Francophone areas do show different socio-political patterns. But how would English foster a stronger deep-set gender bias than French, at least in terms of literacy development? Why would Portuguese have an effect more similar to English than French? A larger sample of countries may clarify whether these differences are real and shed light on why these differences may exist.

7. Broader implications for locating deterministically couched laws

We have found that female and male literacy rates in most countries tested are connected through the relationship $F=M^k$. Bangladesh is the sole major case where some other factor overrides it. The exponent k is found to range from 1.5 to 2.5 for countries with low female literacy around 1980 (all Afro-Asian). This value is unrelated to the UNDP Gender Inequality Index, while possibly affected by the major European language used. When k is constant, the male-female gap in literacy is stubbornly bound to go first up and then down by a well-defined extent. Different policies can only accelerate or slow down the pace at which F and M trace the curve $F=M^k$, from $F=M=0$ to $F=M=1$.

What does the law of female-male literacy say about law-like relationships in social sciences in general, and about ways to locate them? And how could the existence of this particular law be explained?

The broader implications of this study are both substantive and methodological. On the substantive side it offers a specific quantitative relationship between two social characteristics, female and male literacy in a country, over time. This relationship is shown to work for male literacy data ranging from 15 to 90% empirically, and from 0 to 100% by introducing the notion of conceptual anchor points. The impact on male-female literacy gap is automatic. This relationship has a deterministically couched format, akin to many laws in natural sciences. Such relationships are so rare in social sciences that some researchers doubt they can exist. The present result confirms that they can exist, and this should supply an incentive to look for further such relationships. How is one advised to look for them? Now we are shifting to methodology.

The starting point for this study was logical considerations rather than statistical analysis. It made use of the concept of conceptual anchor points stressed by Taagepera (2008: pp. 34-37; 2011). This being so, it would be risky to look for social laws by starting with knee-jerk statistical fits, without immediately imposing logical constraints – in the present case, that the best fit curve must go through 0,0 and 1,1. Once the model $F=M^k$ is confirmed by eyeballing (Figures 1 to 3), should it be further subjected to formal statistical testing? We maintain that little extra could be gained by so doing, when eyeballing is unambiguous, and that much could be lost when statistics are applied inadequately – which often happens. The simplest way to test the degree of fit to $F=M^k$ is to shift to logarithms – $\log F = k \log M$ – and then apply linear regression to $\log F$ and $\log M$. However, if OLS is applied unthinkingly, it would produce a line $\log F = A + k \log M$, which would not respect the conceptual anchor point 1,1 (unless intercept A=0 exactly).

Another difficulty is that, like most law-like expressions in natural sciences, $F=M^k$ can be reversed to $M=F^h$, where $h=1/k$. However, OLS regressions of $\log F$ on $\log M$ and $\log M$ on $\log F$ produce two distinct lines when $R^2<1$. This means that the inverse of the regressed h is not equal to the regressed k, and neither represents the slope of actual trend line – the line that one might draw in as visual best (Flanagan-Hyde, 2006). This problem can be overcome by using symmetric regression (Taagepera, 2008, pp.154-174), but the problems of non-zero intercept A remain. Even if the symmetric regression line is forced to have A=0 (which presents problems of its own), a difficulty remains. Taking logarithms of small fractions (such as M=.145 and F=.032 in the case of Burkina Faso) boosts random error. As a result, the regression line would be unduly affected by the smallest entries. We are not certain how to counteract this effect – but it exists, and we'd rather bypass formal statistics than apply them in a possibly misleading way.

In sum, law-like relationships in social sciences are not likely to be found through statistical methods, and their later testing by statistical means is fraught with pit holes. Besides asking, "How things are?" on the basis of data, one must ask, "How things should be?" on logical grounds.

In the latter respect, the discovery of $F=M^k$ represents a breakthrough but remains incomplete. This equation satisfies logical requirements in terms of anchor points, and fits the data (with one major exception). But why does this relationship come about? How do literate and illiterate males and females interact to produce the empirically observed $F=M^k$? This process of interaction remains to be elucidated. Maybe then we can also answer the question inherent in Figure 1: If both Burkina Faso and Mozambique follow the pattern $F=M^k$, why do they do so with markedly different values of exponent k? Maybe we can then also tell in what way Bangladesh differs from other countries. As one looks at the values of k in Table 1, one observes a rather symmetric distribution around 2.0. (The mean k is 2.04.) One may wonder whether one could treat country-specific values of k as random deviations from 2, so as to make the law $F=M^k$ even more specific: Female literacy in a country increases roughly as the square of male literacy: $F=M^2$. However, what one gains in simplicity of expression one loses in precision. There is nothing special or logical in the simple value k=2.0, in contrast to, say, the exponent 2 for distance in laws of gravitational or electric forces. In the complete absence of gender distinction k=1 would prevail ($F=M$). Protestant countries that insisted on everyone's ability to read the holy script could well have k much less than 1.5. At the other extreme one could imagine cultures even more inimical toward female literacy than what is expressed by k=2.5. Once more, elucidation of the underlying process of interaction could offer answers. The stubborn law of female-male literacy found in our study provides an important starting point for further illumination of the relationship between female and male literacy.

Endnotes

[1] See http://infochangeindia.org/education/news/.

[2] We typed in a variety of phrases into jstor.org and scholar.google.com including "female and male literacy rates", "female literacy rates", "female literacy development", "disparity male and female literacy rates", etc.

[3] See http://huebler.blogspot.com/2007/08/disparity-between-male-and-female.html.

[4] Suppose the overall pattern fits $F=M^{2.0}$, but F has a random error range of $M^{2.0}\pm.01$ (±1 percent point). At M=.50, F=.25, the resulting k can range only from 1.94 to 2.06. But at M=.173, F=.03, k could range from 1.84 to 2.23; and for M=.949, F=.90, from 1.79 to 2.21.

[5] Data retrieved from http://data.uis.unesco.org/Index.aspx?queryid=166.

[6] See http://www.unicef.org/infobycountry/.

[7] There is nothing magical about these two groups. As more countries are added, their k might often be intermediary, around k=2.0. These groupings just offer us a way to graph together countries with similar k values.

[8] See http://infochangeindia.org/women/statistics/

[9] For a discussion of the validity of examining signatures to assess literacy rates see Francois (1977) and Curtis (2007).

[10] We have $G=M-F=M-M^k=M(1-M^{k-1})$. As M increases, $(1-M^{k-1})$ decreases, and a central peak results for G. It can be shown that the peak height is $G=1/k^{1/(k-1)}-1/k^{k/(k-1)}$, reached when $M=1/k^{1/(k-1)}$. For k=1.55, the maximal gap is 0.160, reached at M=.452. For k=2.0, it is 0.25, reached at M=.50. For k=2.52, it is 0.328, reached at M=.544. The gap could also be graphed against F: $G=M-F=F^{1/k}-F=F(F^{1/k-1}-1)$. For k=2.0, the maximal gap 0.25 is reached at F=.25.

[11] Data retrieved from http://hdr.undp.org/en/content/gender-inequality-index-gii.

[12] This duality of OLS lines is well known to statisticians, but few statistics texts for social sciences stress it. Exceptions include Huck and Sandler (1984, pp.60-61), von Eye and Schuster (1998, pp.209-225), and Kennedy (1998, p.141).

References

Curtis, B. (2007) "Beyond Signature Literacy: New Research Directions", *Historical Studies in Education / Revue d'histoire de l'éducation*. 19 (2), 1–12.

Dreze, J., and Loh, J (1995). "Literacy Rates in India and China", *Economic and Political Weekly*, 30(45), 2868-2878.

Flanagan-Hyde, P. (2006) "The Least-Squares Regression Line Is Not a Trend Line. *STATS", The Magazine for Students of Statistics,* 45, 18–20.

Furet, F. (1977) "Lire et écrire: l'alphabétisation des Français de Calvin à Jules Ferry", *Paris: Editions de Minuit*.

Gambell, T. J., and Hunter, D. M. (1999) "Rethinking Gender Differences in Literacy", *Canadian Journal of Education,* 24(1), 1-16.

Huebler, F. (2007) *Disparity between male and female literacy rates* [online]. International Education Statistics. Available from: http://huebler.blogspot.com/2007/08/disparity-between-male-and-female.html [Accessed 12 November 2015].

Huck, S. W., and Sandler, H. M. (2006) *Statistical Illusions: Problems*. New York, NY: Harper & Row.
Infochangeindia. (2004) *India's urban-rural, male female literacy gap closing: report* [online]. Census of India. Available from: http://infochangeindia.org/education/news/. Based on Census of India [Accessed 2 December 2015].
Infochangeindia. (2004) *Changing trends in the gender gap in literacy rates since 1951-2001* [online]. Indo-Asian News Service. Available from: http://infochangeindia.org/women/statistics/ [Accessed 2 December 2015].
Kennedy, P. (1998) *A Guide to Econometrics*. Cambridge, MA: MIT Press.
King, E. M., and M. Hill, A. (eds.). (1993) *Women's Education in Developing Countries: Barriers, Benefits, and Policies*. Baltimore, MD: The Johns Hopkins University Press.
Klein, H. A. (1977) "Cross-cultural studies: What do they tell about sex differences in reading?", *The Reading Teacher*, 30(8), 880-885.
McDougall, L. (2000). "Gender Gap in Literacy in Uttar Pradesh: Questions for Decentralised Educational Planning", *Economic and Political Weekly*, 35(19), 1649-1658.
Magnuson, R. (1985) "Two myths in New France education", *McGill Journal of Education*, 20(3).
Raju, S. (1988) "Female Literacy in India: The Urban Dimension", *Economic and Political Weekly*, 23(44).
Sharma, O. P., and Retherford, R. D. (1987) "Recent literacy trends in India", *New Delhi, IN: Office of the Registrar General & Census Commissioner, India Ministry of Home Affairs*, Govt. of India/New Delhi.
Taagepera, R. (2008) *Making Social Sciences More Scientific: The Need for Predictive Models*, Oxford, UK: Oxford University Press.
Taagepera, R. (2011) "Adding meaning to regression", *European Political Science*, 10, 73-85.
UNICEF. (2015) *Bangladesh* [online]. UNICEF. Available from: http://www.unicef.org/infobycountry/ [Accessed 20 August 2015].
United Nations Development Programme Human Development Reports. (2014) *Gender Inequality Index* [Data File]. Available from: http://hdr.undp.org/en/content/gender-inequality-index-gii [Accessed 20 August 2015].
United Nations Educational, Scientific, and Cultural Organization Institute for Statistics. (2015) *Literacy Rate* [Data File]. Available from: http://data.uis.unesco.org/Index.aspx?queryid=166 [Accessed 1 August 2015].
von Eye, A., and Schuster, C. (1998) *Regression Analysis for Social Sciences*. San Diego, CA: Academic Press

Appendix: Male and Female Literacy Data

Table A. Male (*M*) and female (*F*) literacy, and exponent *k*=log*F*/log*M*.

Countries	Year	Male	Female	k
Afghanistan	1979	0.303	0.050	2.509
Afghanistan	2011	0.454	0.176	2.200
Afghanistan	2015	0.520	0.242	2.170
Bangladesh	1981	0.397	0.180	1.856
Bangladesh	1991	0.443	0.258	1.664
Bangladesh	2001	0.539	0.408	1.451
Bangladesh	2012	0.625	0.551	1.268
Bangladesh	2015	0.646	0.585	1.227
Benin	1979	0.252	0.095	1.708
Benin	1992	0.399	0.166	1.954
Benin	2002	0.479	0.233	1.979
Benin	2006	0.406	0.184	1.878
Benin	2015	0.499	0.273	1.868
Burkina Faso	1975	0.145	0.032	1.782
Burkina Faso	1991	0.196	0.082	1.535
Burkina Faso	1996	0.185	0.081	1.489
Burkina Faso	2003	0.294	0.152	1.539
Burkina Faso	2005	0.314	0.166	1.550
Burkina Faso	2007	0.367	0.216	1.529
Burkina Faso	2015	0.430	0.293	1.455
Burundi	1979	0.343	0.120	1.982
Burundi	1990	0.482	0.275	1.769
Burundi	2000	0.673	0.522	1.642
Burundi	2008	0.888	0.846	1.408
Burundi	2015	0.882	0.831	1.474

CentralAfricanRepublic	1975	0.296	0.084	2.035
CentralAfricanRepublic	1988	0.480	0.203	2.173
CentralAfricanRepublic	2000	0.668	0.353	2.581
CentralAfricanRepublic	2010	0.507	0.244	2.077
CentralAfricanRepublic	2015	0.507	0.244	2.077
Egypt	1976	0.536	0.224	2.399
Egypt	1986	0.570	0.314	2.061
Egypt	1996	0.672	0.436	2.088
Egypt	2005	0.830	0.594	2.795
Egypt	2006	0.746	0.578	1.871
Egypt	2010	0.803	0.635	2.070
Egypt	2012	0.817	0.658	2.071
Egypt	2015	0.822	0.654	2.166
Guinea-Bissau	1979	0.333	0.086	2.231
Guinea-Bissau	2000	0.576	0.275	2.340
Guinea-Bissau	2012	0.698	0.439	2.290
Guinea-Bissau	2015	0.718	0.483	2.197
Liberia	1984	0.449	0.193	2.054
Liberia	1994	0.544	0.313	1.908
Liberia	2004	0.608	0.257	2.731
Liberia	2007	0.608	0.270	2.631
Liberia	2015	0.624	0.328	2.364
Mali	1976	0.135	0.057	1.431
Mali	1998	0.267	0.119	1.612
Mali	2003	0.327	0.159	1.645
Mali	2006	0.349	0.182	1.618
Mali	2010	0.434	0.203	1.910
Mali	2011	0.433	0.246	1.676
Mali	2015	0.482	0.292	1.687
Morocco	1982	0.437	0.175	2.105
Morocco	1994	0.552	0.287	2.101
Morocco	2004	0.657	0.396	2.205
Morocco	2008	0.684	0.426	2.247
Morocco	2009	0.689	0.439	2.210
Morocco	2011	0.761	0.576	2.020
Morocco	2015	0.786	0.588	2.205
Mozambique	1980	0.440	0.121	2.572
Mozambique	1997	0.548	0.250	2.305
Mozambique	2003	0.656	0.332	2.615
Mozambique	2009	0.674	0.365	2.555
Mozambique	2015	0.733	0.454	2.542
Nepal	1981	0.317	0.092	2.077
Nepal	1991	0.492	0.174	2.465
Nepal	2001	0.627	0.349	2.255
Nepal	2011	0.711	0.467	2.232
Nepal	2015	0.764	0.531	2.351
Pakistan	1981	0.354	0.148	1.840
Pakistan	1998	0.553	0.290	2.090
Pakistan	2005	0.641	0.354	2.335
Pakistan	2006	0.677	0.396	2.375
Pakistan	2008	0.689	0.401	2.453
Pakistan	2009	0.686	0.403	2.411
Pakistan	2010	0.689	0.410	2.393
Pakistan	2011	0.670	0.420	2.166
Pakistan	2015	0.695	0.458	2.146
Senegal	1988	0.369	0.179	1.726
Senegal	2002	0.511	0.292	1.834
Senegal	2006	0.523	0.330	1.710
Senegal	2009	0.618	0.387	1.973
Senegal	2011	0.663	0.404	2.205
Senegal	2015	0.697	0.466	2.115

Data from UNESCO Institute for Statistics Literacy Rates (2015).