I got 99 problems
but being female ain't one…Or is it?

Estimating the Gender Gap on Math Test Scores

Alexis Cerda, Arizona State University
Angelica Meinhofer, University of Puerto Rico
August 14, 2009

I. Introduction

There have been several studies about the gender gap in math test scores (Felson, 1990; Fan, 1997; Leahey, 2001). Since mathematics achievement plays an important role in student's subsequent schooling, career choices, and professional achievement it is imperative to study the gender difference in mathematics achievement. Women receiving PhDs in the field of mathematical sciences has risen steadily from around the low 20% in the early 1990s to almost 30% in the 2003-2004 (source). Although we have seen an increase of women in the fields involving mathematics, the fact percentages are still low and women are still underrepresented. For example, in 1999-2000, among the recipient bachelor's degrees in the United States only 2% of women majored in engineering compared to 12% of men (Zafar, 2009). It is important to note that no woman has yet to win the Fields Medal or the Nobel Prize in economics.

One leading explanation for women being underrepresented in mathematical sciences is that males perform better in math than females. It is unclear as to why women's mathematical achievement is below men's mathematical achievement. However, if we can gain better understanding of when the gender gap in math appears and the magnitude, it may be a critical factor for advancing women in careers involving mathematics.

We look closely at the study done by Fryer and Levitt (2004) *Understanding The Black-White Test Score Gap In The First Two Years Of School* to gain insight on how to proceed in our own research. The Fryer and Levitt study focuses on the black-white test score gap among incoming kindergartners. We expand their research by controlling for similar observable characteristics over multiple age groups and focus on the gender test score gap. In this paper we utilize the National Longitude Surveys to supply new information on the gender gap. We are particularly interested in gender differences at the high end and low end of the score distribution.

We analyzed two groups of students around the age they enter school (age 4.5 to 6 years old) and students who are in the last two years of middle school (11.5 to 13 years old). Within those two age groups we examine students contained in two extreme ranges of the math score distribution: those who scored within the top quartile (at or above the 75th percentile), and those who scored in the bottom quartile (at or below the 25th percentile).

We believe it is important to analyze the top and bottom percentiles since the top percentile may early identify the mathematically talented females who have a higher probability of later entering mathematically fields. We would like to analyze how top scoring female compare to their male counterparts and how their performance change as children age.

The format of this paper is as follows: section II summarizes the review of the literature, section III provides a description of our data set, section IV, V, and IV presents the empirical estimation and our results, section VII concludes our paper and section VIII is our appendix with all the tables.

II. Literature Review

There has been a lack of consistency in the literature as to how early the gender gap begins and the causes of the gender gap. Studies as earlier as 1957 reports differences in problem solving skills can be partly accounted for differences in sex-roles identification (Milton). Since feminine roles are typically more verbal and less quantitative task orientated and masculine roles tend to be more quantitative task orientated, imitating same sex parent will result in differences in ability between boys and girls (Milton,1957). Others claimed females' attitudes towards math and negative stereotypes may have influenced their participation in math classes. Mattel in the 1990s, marketed a talking Barbie doll whose key phrase was “Math is hard”(Felson ,1991). However, support for this claim has been inconclusive and inconsistent (Felson , 1991). For instance, females earn higher grades than males in most subjects including math. However on standardize tests such as the Scholastic Aptitude Test (SAT) on average males perform better than females.

Males are not only out performing females on the SAT math section, the American College Test (ACT) entrance exam as well. The gender difference on the two college entrance exams was on average 1.2 points in 1996 (Hyde, 1981). A male advantage in math has been found in the general population (Hyde, 1981; Leahy, 2001; Fan, 1997). Though the advantage seems to be limited to specific type of mathematical skills. Several studies have claimed males having greater spatial-visual ability than females cause gender differences in math. For instance, exams testing analytic spatial-visualization ability and quantitative reasoning ability have shown significant math achievement in favor male students (Battista, 1990).

Recent studies have focus less on explaining why the gender gap exists and more on when the gap appears and the magnitude of the gap. One study asserts that a slight male advantage in mathematics appears in the later years of high school and specifically in geometry (Leahey and Guo, 2001). Leahey and Gou did not find significant gender differences in mathematics performances at early grade levels. Fan and Chen looked at gender differences in the high-end score distribution, however; their research is limited to only looking at descriptive statistics such as the count of females in the high-end distribution and mean test score comparisons. Among students who scored above the 95^{th} percentile in 12^{th} grade, males outnumbered females by a 2:1 ratio. They also found a pattern of gender differences for the total sample at the extreme score ranges in Asian, Hispanic, and White sample groups. However, African-Americans had a larger percentage a females in the extreme score ranges (Fan and Chen, 1997).

III. Data

The data set used in this study is gathered from the National Longitudinal Surveys, which the Bureau of Labor Statistics designed. Since our question of interest targets the gender gap in math test scores as children age, we were particularly interested in two of the surveys: NLSY79 and NLSY79 Child and Young Adult. The NLSY79 is a nationally representative sample of 12,686 young men and women who were between 14 to 22 years old in which they were first surveyed in 1979. These individuals were interviewed annually through 1994 and afterwards were interviewed biannually until 2006. The NLSY79 Child and Young Adult survey was administered biannually (starting in 1986) to all age appropriate children born to NLSY79 female respondents. The availability of longitudinal child information on variables that include, but are not limited to birth weight, race, test scores, and economic well-being and marital status of their NLSY79 mother, allowed us to examine the link between mothers and the development of their respective children.

For this study, data was organized as a cross section in which children ages were held fixed at different points in time (biannually from1986 to 2006). Using Fryer and Levitt's *Understanding the black-white test score gap in the first two years of school* (2004) as our starting point (base case), we extended their study by regressing the effect of multiple covariates on math test score percentile as children aged for seven different age ranging 18 months apart: 4.5-6 years (54-72 months); 6-7.5 years (72-90 months); 7.5-9 years (90-108 months); 9-10.5 years (108-126 months); 10.5-12 years (126-144 months); 12-13.5 years (144-162 months); 13.5-15 years (162-180 months). Since numerous studies conclude that females outperform males at early years of their elementary school education, but with time are surpassed by the latter (Guo and Leahey, 2001), we found it appropriate to track the changes of the coefficients for female students as children aged. To further understand the magnitude of this gender gap, we decided to explore what happens at the top and bottom 25 percentile math scores (extended case). For the base case, the dependent variable is the Peabody Individual Achievement Test (PIAT) math percentile obtained by each age group. For the extended case, the dependent variable is binary (0-1 indicator) and it reflects the outcome (1) in which a student scored in the top and bottom quartiles of the PIAT math test. Since children were tested at different points in time, different observations for the same individuals will be included more than once (at the different age groups).

Summary statistics for part of the base case and extended case variables used are available on tables 1 and 2, respectively. Other dummy covariates such as year effects, order of birth and number of siblings were included in the regressions, but are not included in the tables.

Table 1 shows math percentile averages across age groups for the entire sample and for individual case of males and females. On average, females slightly outperform males for the first two age groups (4.5 – 6 and 6 – 7.5), but for the group that includes the ages between 7.5 – 9, we begin to see males surpass females by a minor but relatively constant gap. Table 1, demonstrates that NLSY79 oversamples minority groups; blacks represent 27.67% of the sample, while Hispanics represent 19%. This significantly deviates from true population in the US: Hispanics and blacks account approximately for 15% and 12.8% percent respectively. The mother's in the sample are on average 25 years old when they have their first child. Also, the average amount of education completed highest 12.60, which implies a high school diploma.

In Table 2 it can be observed that in the top quartile when children are between the age 4.5 – 6, there is a proportional female-male presence of 51% and 49% respectively. However, this is not the same at the bottom quartile, in which females account for 42% and males represent 58%, a 16% difference. On the other hand, when children reach the age 12 – 13.5, males represent 57% of the top quartile, while females now account for 43% (there are more male students at the top than female students). For the bottom 25 percentile, females and males represent 49% and 51% of the sample (females and males are equally likely to be found in the bottom). Interestingly but perhaps expected, children who score at the bottom have mothers that on average had their first child at a younger age than mothers of children in the top. A parallel story occurs for mother's education, which on average is higher for children at the top relative to children at the bottom.

IV. Estimating Gender Test Score Gaps

Table 4 presents a series of estimates with our dependent variable as the math percentile for the test taken at the ages 4.5-6 years old, 6-7.5 years old, 7.5-9 years old, 9-10.5 years old, 10.5-12 years old, 12-13.5 years old and 13.5-15 years old. The specifications estimated are of the form

$$\mathbf{Math\_Percentile}_{ij} = \beta_0 + \beta_1 \mathbf{Female}_i + \beta' \mathbf{X}'_{ij} + \beta\,' \mathbf{H}'_i + \varepsilon_{ij}$$

where *i* indexes students and *j* indexes each age category. The coefficient on female captures the gap between male and female students for each age group. The vector of covariates denoted $X_{ij,}$ includes the regressors that change with age groups (for age specific years). The variables in X_{ij} are age, household's net income and mother's marital status. Household's net income (in 2006 dollars) tries to capture socioeconomic status for each observation at different points in time, while mothers marital status (at each age) should somewhat captured the house environment (though this is not a precise measurement nor does it apply equally to the different minority groups). It was coded as one if the mother was married and the spouse was present, zero otherwise.

The vector H_i consists of regressors that are age group invariant. The variables in H_i includes mothers age at time of first birth, birth weight, two indicator variables for mother education; whether the mother dropped out of school, or mother has at least one year of college and the excluded category is if the mother is a high school graduate. The other variables in included are indicator variables for welfare, Hispanic, black (white is the excluded category), number of siblings, order of birth and year effects. Mothers age at time of first birth, birth weight, and mother's education and welfare attempts to capture socioeconomic status. It is worth noting that welfare is coded as one if the household ever (1986-2006) received welfare, zero otherwise. This could be misleading since several households could have only received welfare compensation for a short period of time. However, in situations like this we believe that these households would still be consider low income even though they receive welfare for small period of time. Although gender is random (e.g. females and males are equally likely to be found in different racial groups, income distributions, etc.), we control for race by including indicator variables. We tried to include variables that would control to some extent for the house environment; therefore we included the number of siblings and order of birth. We converted them into binary to avoid forcing a linear relationship, since the observations for these variables are mostly going to cluster around certain values.

Originally we weighted the sample to account for the over sampling of minorities, yet again, when we compared our estimates from both, the weight and unweighted, there was little change. Thus, we have decided to only present unweighted results.

Base Case

Table 3 illustrates the estimated coefficients of a regression for each age group with the where the indicator variable for female is the only regressor. We found that for the age groups 4.5-6 and 6-7.5, female students outperform male students in the PIAT math section. Furthermore, the coefficients are significant at the 95% level for both cases. This advantage is apparent in the first age group, for which being a female is associated with scoring 3.43 percentile points higher than males. As children age, however, the coefficients reverse in favor of males. The negative coefficients are statistically significant for the age groups 10.5-12, 12-13.5 and 13.5-15. We see that the gender gap continue to increase as children age (it reduces at the age 13.5-15, however fewer observations are available for this particular age group). By age 12-13.5 female students end up with a disadvantage of 2.68 percentile points below male students.

We decided to add the covariates mentioned earlier to explore what would the change in the gender gap. We found minor changes in the female coefficient, which is explained by the fact that gender is a random and independent variable. As found with single regressor model estimates, we continue to observe that in the early age groups, females are outperform males (3.58 percentile points at age 4.5-6) but as children age, males surpass females, who end up with a disadvantage of 2.34 percentile points at age 13.5-15. Looking at the race indicator variables, we see find large gaps for blacks (e.g. -13.52 at age 12-13.5) and Hispanics (e.g. -10.94 at age 10.5-12) relative to the omitted variable whites. Putting things into perspective, it seems minority groups are in a more disadvantageous position than females. Regardless of, the gender gap is still an important problem that should be investigated, since should be no reason for which on average, males outperform girls of similar characteristics.

Other interesting coefficients are the indicator variables for mother's education. Looking the age 10.5-12 group for example, we observe that children whose mother didn't complete high school score 4.49 percentile points below the students whose mother did finish high school. On the other hand, children whose mother pursued higher education, score 5.66 percentile points above the students whose mothers completed high school. Mother's marital status was not found to be statistically significant at any age group.

V. Estimating The Probability Females Are In The Top And Bottom Quartile Using A Linear Probability Model

Table 5 reports the estimated coefficients for the probability that female students will be in the top or bottom quartile for ages 4.5-6 years old. The model estimates are of the form

$$\mathbf{TopQuartile6}_i = \beta_0 + \beta_{1i}\ \mathbf{Female}_i + \beta' \mathbf{X}'_{ij} + \beta\ '\mathbf{H}'_i + \varepsilon_i$$

where TopQuartile6$_i$ is binary, *i* indexes the students and *j* indexes the age category , so that

$$\Pr(\text{TopQuartile6}_i = 1 \mid \text{Female}_i, X'_{ij}) = \beta_0 + \beta_{1i}\ \text{Female} + \beta' X'_i + \varepsilon_i$$

Female is an indicator variable where it equals one when the student is a female and zero otherwise. TopQuartile6 is indicator variable as well, where it equals one if the student's math score was at or above 75th quartile and zero otherwise.

Table 5 also reports the estimated coefficients for the bottom quartile for children between the ages of 4.5-6 years old. The model for the bottom quartile is of the same form

$$\mathbf{BotQuartile6}_i = \beta_0 + \beta_{1i}\ \mathbf{Female} + \beta' \mathbf{X}'_{ij} + \beta\ '\mathbf{H}'_i + \varepsilon_i.$$

where BotQuartile6$_i$ is an indicator variable, and it takes the value of one if the student is at or below the 25th percentile. Listed below are the models for the top and bottom quartiles for children between the ages of 12-13.5, respectively

$$\mathbf{TopQuartile13}_i = \beta_0 + \beta_{1i}\ \mathbf{Female} + \beta' \mathbf{X}'_{ij} + \beta\ '\mathbf{H}'_i + \varepsilon$$

$$\textbf{BotQuartile13}_i = \beta_0 + \beta_{1i}\ \textbf{Female} + \beta'\mathbf{X}'_{ij} + \beta\ '\mathbf{H}'_i + \varepsilon_i.$$

In table 5, we can see that the estimated coefficient for female is .032 for the top quartile for children who between the ages of 4.5-6 years old. This implies that the probability of a female student being in the top quartile is 3.2% higher than male student with a p-value of 0.024 and a z-score of 2.26. Similarly, female students are estimated to be 7% less likely in the bottom quartile with a p-value of 0 and a z-score of 4.76. For children between the ages of 12-13.5 females are predicted to be in the top quartile with a probability of 5.7% less than male students with a p-value of 0 and a z-score of -4.61. Female students are predicted to be in the bottom quartile with equal probability as male students, since the relationship between the bottom quartile and being female is not significant.

Although we can account for heteroskedasticity by using robust OLS standard errors, some may be concern that some of the fitted values are out outside of the zero-one range. Therefore we estimate probability females are top and bottom quartile using a different model to compare results.

VI. Estimating the probability females are in the top and bottom quartile using a Probit Model

Table 6 presents the point estimates for the probability that female students will be in the top or bottom quartile for ages 4.5-6 years old. The estimated probabilities are of the form

$$\textbf{Pr(TopQuartile6}_i \mid \textbf{Female}_i, \mathbf{X}'_{ij}) = \Phi(\ \beta_0 + \beta_{1i}\ \textbf{Female} + \beta'\mathbf{X}'_{ij} + \beta\ '\mathbf{H}'_i\ ) + \varepsilon_i$$

where *i* indexes students, j indexes age categories, and Φ is the cumulative standard normal distribution function. Female and TopQuartile6 are the same indicator variables in the linear probability model. The vector of other covariates in included specifications, denoted X_i, varies across columns in table 6.

Columns in table 6 report the point estimates for the bottom quartile for children between the ages of 4.5-6 years old. The estimated probabilities are of the form

$$\mathbf{Pr(BotQuartile6_{\mathit{i}}\mid Female_{\mathit{i}}, X'_{\mathit{ij}}) = \Phi(\,\beta_0 + \beta_{1\mathit{i}}\, Female + \beta' X'_{\mathit{ij}} + \beta\,' H'_{\mathit{i}}\,) + \varepsilon_{\mathit{i}}.}$$

This specification is exactly the same as before expect the dependant variable, BotQuartile6$_i$ is an indicator variable, where it takes the value of one if the student is at or below the 25th percentile.

The specification for the top and bottom quartile for students between the ages 12-13.5 years old is listed below, respectively

$$\mathbf{Pr(TopQuartile13_{\mathit{i}}\mid Female_{\mathit{i}}, X'_{\mathit{ij}}) = \Phi(\,\beta_0 + \beta_{1\mathit{i}}\, Female + \beta' X'_{\mathit{ij}} + \beta\,' H'_{\mathit{i}}\,) + \varepsilon_{\mathit{i}}}$$

$$\mathbf{Pr(BotQuartile13_{\mathit{i}}\mid Female_{\mathit{i}}, X'_{\mathit{i}}) = \Phi(\,\beta_0 + \beta_{1\mathit{i}}\, Female + \beta' X'_{\mathit{i}} + \beta\,' H'_{\mathit{i}}) + \varepsilon_{\mathit{i}}}$$

where the independent variables and the dependent variable have same meaning as the linear probability models for students in the top and bottom quartile.

The probit coefficients do not have simple interpretations. This model is best interpreted by computing predicted probabilities and the effect of a change in one of our regressors. For that reason we report the marginal of effects of each regressor in table 7. Which means the change in the probability for an infinitesimal change in each independent, continuous variable and a discrete change in probability for indicator variables. The marginal effects are separately evaluated at the mean scores of top and bottom quartile.

The coefficient in table 7 for female is positively related to the probability of being in the top quartile for students between the ages 4.5-6 years old. In fact, when children are around the age of 6, females are expected to be 7% less likely to be in the bottom quartile and 3% more likely to be in top quartile with p-values of 0 and .013 and a z-score of –4.69 and 2.48

respectively. In table 7 we can now see that being female is negatively related to the probability of being in the top quartile for students around the age of 13. Females are predicted to be about 6% less likely to be in the top quartile with a p-value of 0 and a z-score of –4.51. The relationship between being female and the probability of being in the bottom quartile is not statistically significant (p-value=.5 and z=-.68) which suggest females are estimated to be equally likely in the bottom quartile around the age 13.

One might be concern that assuming the errors are normally distributed is incorrect. However, the probit estimator is consistent and normally distributed in large samples. There are advantages and disadvantages for using a probit and linear probability model and we could not distinguished which model was better. Hence, we report estimates for both models though the coefficient estimates are very similar. As mentioned before we did not report the estimates for when the model was weighted.

VII. Conclusion

We find that for children around the ages of 4.5-7.5, being a female is associated with scoring at a higher percentile relative to males. As children age, specially starting at age 10, males begin to gain advantage over girls in the math section of the PIAT. It is clear that for this standardize test, a gap exists in favor of males. In fact, when children are around the age of 6, females are expected to be 7% less likely to be in the bottom quartile and 3% more likely to be in top quartile. Females are predicted to be about 6% less likely to be in the top quartile and equally likely in the bottom quartile around the age 13. Yet again, we can’t say this gap is very large, especially relative to the gap of minority groups. What is most important is to understand the reason for this gap and why females are still underrepresented in mathematical related fields. The answer to that question is clearly outside the scope of this study, but we do find results that suggest it begins in elementary school and continues as children age. We think further research should include a nationally representative longitudinal study following the same students through time and tracking how the gender gap changes for the top quartile from age 13 until the end of high school. As well as tracking males and females at top and bottom quartiles in order to follow their further development and test performance. Do they stay at the top or bottom percentile through time? Perhaps then there will be enough information to make any policy recommendations.

Works Cited

Battista, M. (1990), “Spatial visualization and gender differences in high school geometry”, *Journal for Mathematics Education*, 21, 47-60

Fan, Xitao & Chen, Michael (1997), “Gender differences in mathematics achievement: Finding from the National Education Longitudinal Study of 1988”, *Journal of Experimental Education*, 65, 229-243

Felson, R., & Trudeau, L. (1991), “Gender Differences in Mathematics Performance”, *Social Psychology Quarterly*, 54, 113-126

Leahey, E., & Guo, G. (2001), “Gender Differences in Mathematical Trajectories”, *Social Forces*, 80, 713-732

Milton , G. A. (1957), “The Effects of Sex-Role Identification Upon Problem-Solving Skills”, *Journal of Abnormal and Social Psychology*, 55, 208-212

Zafar, Basit (2009), “College Major Choice and the Gender Gap”, *Federal Reserve Bank Of New York Staff Reports*, no. 364