A New Perspective on Stereotypical Gender Differences in Test Scores*

Devin G. Pope
The Wharton School
University of Pennsylvania

Justin R. Sydnor
Department of Economics
Weatherhead School of Management
Case Western Reserve University

This Draft: December, 2008

Abstract

The causes and consequences of gender disparities in standardized test scores – especially in the high tails of achievement – have been a topic of heated debate. The existing evidence on standardized test scores largely confirms the prevailing stereotypes that more men than women excel in math and science while more women than men excel in tests of language and reading. We provide a new perspective on this gender gap in test scores by analyzing the variation in these disparities across geographic areas. We illustrate that male-female ratios of students scoring in the high ranges of standardized tests vary significantly across the U.S. This variation is systematic in several important ways. In particular, states where males are highly overrepresented in the top math and science scores also tend to see women highly overrepresented in top reading scores. This pattern suggests that states vary in their adherence to stereotypical gender performance, rather than favoring one sex over the other across all subjects. Since biological differences between the sexes are unlikely to vary at the state level, while cultural and education environments do, the variation we find speaks to the nature-vs.-nurture debates surrounding test scores and suggests environmental differences in the U.S. greatly impact gender disparities in test scores.

Contact Pope at dpope@econ.berkeley.edu and Sydnor at justin.sydnor@case.edu.

*We thank Betsey Stevenson, Uri Simonsohn, and seminar participants at Case Western Reserve University, the University of Pennsylvania and Stanford University for helpful comments and suggestions. All errors are our own.
Because of the role they play in discussions of gender disparities in economic outcomes, gender differences in standardized test scores in the United States have long been a contentious issue. The topic was never more controversial than in the aftermath of Larry Summers’ speech at an NBER Conference in January, 2005. In discussing the underrepresentation of women in tenured positions in science and engineering at top universities, Summers suggested that one hypothesis for these patterns was the possibility of “different availability of aptitude at the high end” in math and science between men and women. These comments sparked an intense debate on the nature vs. nurture component of sex differences in cognitive abilities.

Whatever one’s opinion of his comments, Summers’ discussion and the arguments that ensued were clearly motivated in part by the existing research on gender differences in test scores. This literature has largely confirmed the stereotype that males excel on tests of math and science, while females excel on tests of language and reading. More precisely, although males and females exhibit similar average scores on most tests, due to sex differences in the variance of test scores and slight differences in means, significantly more males than females score in the very high ranges on science and math tests and significantly more females score highly on language and reading tests (Hedges and Nowell (1995), Hussein and Millimet (2008), Hyde et al. (2008)). For example, Hedges and Nowell (1995) review data from six national studies conducted between 1960 and 1992 and find consistent gender differences in the tails of the test-score distributions. Across the six math tests they examine, the male-female ratio for students scoring in the 95th percentile of the national distribution ranges from 1.50 to 2.34 and is above 2.0 for half of the tests. Taken together the existing research suggests large disparities in gender ratios at the upper tail of the test-score distribution in the United States that have been pervasive over time and across a variety of tests.

1 The full transcript of Summers’ comments is here: http://www.president.harvard.edu/speeches/2005/nber.html.
2 For a good example of the debate surrounding Summers’ remarks, see the transcripts of the public debate on the topic between Harvard psychologist Elizabeth Spelke and Steven Pinker here: http://www.edge.org/3rd_culture/debate05/debate05_index.html.
This paper offers a new perspective on the pervasiveness of gender differences in test scores. We examine the degree of state-level geographic variation in gender disparities on standardized test scores in the United States. Using individual-level data on math, science, and reading tests given to 8th graders since 2000 through the National Assessment of Educational Progress (NAEP), we compute gender ratios in the upper tails (95th and 75th percentiles) of the test-score distribution for each state. We are able to replicate the standard gender disparity findings at the national level, but these national patterns hide large and statistically significant variations in gender gaps at the state level. The sex differences on test scores in the most gender-equal states are less than half the size of the sex differences that are found in the most gender-unequal states. There is more variation in state-level gender gaps than one would expect to find randomly, and this variation is geographically clustered. For example, in the New England census division (CT, NH, MA, ME, RI, and VT), the ratio of males to females scoring above the 95th percentile on the science and math tests are 1.46 and 1.29, respectively, while in the East South Central census division (AL, KY, MS, and TN) the male-female ratios are 2.14 and 1.57.

Perhaps more surprising than the significant geographic variation and clustering is that states which have smaller gender disparities in stereotypically male-dominated tests of math and science also tend to have smaller disparities in stereotypically female-dominated tests of reading. For example, the New England census division, which has the lowest male-female ratios in the 95th percentile on math and science also has the lowest female-male ratio (2.067) at the 95th percentile on the reading test. Thus, the variation across states in test-score disparities is not simply a reflection of some states improving the performance of females relative to males. Rather some states appear to be more gender equal across all tests and adhere less to gender stereotypes in both directions.

These results show that adherence to stereotypical gender norms on standardized tests varies systematically at the state level. From a policy standpoint, this is important because it highlights that...
gender disparities are not immutable. The results also suggest where to look for both models of gender equality and for regions within the United States where gender stereotypes are a bigger issue.

From a more theoretical standpoint, these results also speak to the nature-nurture debate surrounding cognitive ability and test scores. It is unlikely that biological distinctions between males and females that differ at the state level can explain the patterns that we find. As such, it seems reasonable to interpret the variation we find as the effect of different environmental forces. Furthermore, since there is much about the social and educational environment within the United States that does not vary at the state level, the variation we observe is likely a lower bound on the effect of different environments on gender ratios in high-end test-score performance. The variation does not completely explain stereotypical gender performance – even the most gender-equal states display ratios in the direction of prevailing stereotypes – which leaves room for a partial biological/genetic root to test-score differences. Yet there is clearly enough existing variation in the gender ratios across states to suggest that environment plays a very large role in these sex differences in high performance.

A natural question is whether we can identify the environmental factors that affect gender gaps in test scores at the high end. It is clearly difficult to establish causal mechanisms for these state-level patterns, but we investigate some of the correlates of test-score gender equality at the state level. Perhaps not surprisingly, given the discussion of census regions above, we find that states with larger gender disparities have lower median income and less educated adult populations. We also investigate state-level culture more directly using evidence on attitudes toward gender issues from the General Social Survey and an early NAEP test wave. We focus on questions about

---

3 This paper adds to the literature which addresses the question of nature vs. nurture across a wide array of economic outcomes (see Sacerdote (2008) for a review). This paper also adds to the small economics literature which discusses biological and environmental impacts on test scores. Specifically, Guiso et al. (2008) find that the gender gaps in test scores (in all subjects) are higher in some countries than in others and Figlio (2008) provides evidence that girls with more feminine names – even when looking within families – are less likely to select into math and science courses than their counterparts.
attitudes toward abortion and whether it is preferable for women to work in the home rather than in the labor market. We also use a measure of gender equality from earlier versions of the NAEP tests that asked students whether they think that “math is for boys”. These gender-attitude measures are very highly correlated with adherence to test-score stereotypes at the state level even when controlling for individual-level demographic characteristics for the GSS questions. In fact at the 95th percentile, state-level differences in attitudes toward women working at home can explain 44% of the variation in state-level gender disparities. Although these results do not isolate clear causal mechanisms for gender differences in test scores, they suggest that cultural gender attitudes play an important role in stereotypical gender performance in test scores.

The remainder of the paper proceeds as follows. Section I describes the data used in this analysis and the results and Section II provides a brief discussion and conclusion.

I. Data and Results

Data. The test-score data for this study come from the National Assessment of Educational Progress (NAEP), which administers standardized tests using representative probability samples at both the national and state levels. The NAEP is arguably the best source of standardized test-score data for making state-level comparisons in the United States.

We analyze individual-level data from the state NAEP tests in science, math, and reading given to 8th graders in 2000, 2003, and 2005. In order to circumvent questions regarding differences in gender disparities across races, which may correlate with geographic areas, we focus exclusively on white students. The sample size for minority students is too small to obtain inferences for non-

---

4 Throughout the analysis we use the individual-level sampling weights provided in the NAEP data.
5 The data were made available to us through a restricted-access license. Test-score information is available at the state-level for the math test in 2000, 2003, and 2005, the science test in 2000 and 2005, and the reading test in 2003 and 2005.
white races in each state. Pooling across years, there are (142,121), (251,867), and (190,710) usable observations for the science, math and reading tests, respectively.

**National Patterns.** Aggregated nationally, these data confirm the established patterns of gender differences on tests scores. There are only slight differences in the mean scores in favor of males on the math and science tests and somewhat more strongly in favor of females on the reading test. However, as Figure 1 reveals, there are substantial differences in the gender ratios of students scoring in the higher percentiles on these tests. The ratio of males to females scoring in the top 25 percent is 1.33 for science and 1.17 for math, and rises in the top 5 percent to 1.87 for science and 1.40. The disparities in favor of women on the reading test are even stronger, with a female-male ratio of 1.62 in the top 25 percent and of 2.31 in the top 5 percent.

**State-Level Variation in Gender Ratios.** We examine variation at the state level in the gender ratios in the upper tails of the test-score distributions. To do this, we compute a top 25% and a top 5% cutoff value for each test subject and year using a cutoff based on the national sample of test scores. At each of these two cutoff points, we calculate the ratio of the number of males to females scoring above the cutoff in each state.

With one exception, the prevailing stereotypes show up in all states at both the 95\textsuperscript{th} and 75\textsuperscript{th} percentiles. At the 95\textsuperscript{th} percentile, the two smallest male-female ratios (i.e., most gender equal) in math are 0.81 (HI) and 1.06 (NY) and in science 1.30 (MA) and 1.43 (WA). The two smallest female-male ratio in reading are 1.75 (MA) and 1.88 (RI).

There is however, a large amount of variation in these ratios. For instance, in contrast to the low ratios at the 95\textsuperscript{th} percentile on the math test for Hawaii and New York, the two highest ratios

---

\[6\text{ The standardized mean difference } d \text{ (males – females) is 0.17 for science, 0.06 for math, and -0.38 for reading.}\]
\[7\text{ The sample sizes in the NAEP are not large enough to extend the analysis to the county or MSA level.}\]
\[8\text{ Ideally we would conduct these tests even higher in the distribution, say the top 0.1\%, since the debate about gender differences in ability often surrounds the very extreme levels of performance. However, there simply is not enough power, even in the NAEP data, to study state-level variation at those extreme tails.}\]
\[9\text{ The one exception is Hawaii, which has a male-female ratio of 0.81 on the math test at the 95\textsuperscript{th} percentile.}\]
are roughly twice as high – 1.93 for Oklahoma and 2.07 for Kentucky. On the science test, three states (UT, MS, NJ) have male-female ratios above 3.0 – more than twice the low-end ratios observed in Massachusetts and Washington. On the reading test the highest female-male ratio at the 95th percentile occurs in Utah at a staggering 4.47, implying that 82% of the Utah students scoring at the top of the reading test were female.

Of course, ex-post we would expect to find some level of variation in these gender ratios. In order to statistically test for excess variation in these gender ratios (above what one would expect due to random chance) an F-test can be used to test the null hypothesis that these gender ratios are the same across states. For each test at both the 95th and 75th percentiles the F-test rejects the null hypothesis that the gender ratios are the same across states, with p-values below 0.05 in every case.

An important pattern emerges when we examine the correlation of the gender ratios on the three tests. There is a correlation between the gender gaps on math and science at the state level. For instance, states like Utah and Oklahoma have high male-female ratios on the 95th percentile of both the math and science tests – Utah ranks 45th on the math and 48th on the science, while Oklahoma ranks 49th on the math and 46th on the science. Initially, one might suspect, then, that these states have environments that favor boys over girls. Yet looking at the reading tests, we see that in these same states girls outperform boys on the reading tests by an unusually large margin. Utah has the highest female-male ratio at the 95th percentile on the reading test and Oklahoma has the 8th highest. In fact, this pattern holds broadly in our data, as Figure 2 illustrates. This figure graphs the average male-female ratio in the 95th (Panel A) and 75th (Panel B) percentiles of the math and science tests against the female-male ratio at these same cutoffs on the reading test for each state. There is a strong correlation at both percentiles (P < .01) – states where boys outperform girls in math and

---

10 Similarly, a chi-squared test can be used to analyze whether the mean gender gaps are different across states. However, the F-test allows us to include the sampling weights provided in the NAEP data, which a standard chi-squared test does not.
science at the high end of the distribution also tend to see girls outperform boys in reading. This pattern suggests that states adhere more or less strongly to the prevailing gender stereotypes in test performance, rather than simply favoring one sex over the other.

States with high levels of stereotypical gender differences in test scores also appear to be clustered geographically. To examine this geographic clustering, we create a state-level “stereotype adherence index” (SAI) at a given cutoff by averaging a state’s male-female ratio in math and science with the state’s female-male ratio in reading. Figure 3 presents a map of the United States shaded to represent the level of the SAI for each state for the 95th (Panel A) and 75th (Panel B) percentiles. The states with a very high SAI (i.e., large gender disparities) are predominately found in the South and Mountain West, while the states with a very low SAI are mostly found in the West, Southwest, and Northeast.

Of course, obvious problems of potential bias arise when analyzing ex-post groupings of states in this way, and it would be preferable to test whether regions vary based on a-priori groupings. Census divisions are useful ex-ante geographic state groupings for this analysis. There are large disparities in the SAI by census division. For example, the SAI in New England (1.61), which has the lowest SAI, is 24% lower than the highest SAI (2.12) found in the East South Central division (AL, KY, MS, and TN). Looking at all of the census divisions, an F-test of the equality of the SAI across divisions rejects with p-values below 0.01 at both the 95th and 75th percentiles.

**Correlates with Stereotypical Gender Disparities.** Although it is difficult to establish causal mechanisms for these state-level variations – especially given the potential relevance of hard-to-measure characteristics like culture and gender attitudes – it seems natural to investigate the state-level characteristics that correlate with stereotypical test-score gender disparities. Understanding these correlations may help focus policymakers on the areas with greatest gender disparities and will
hopefully provide directions for future research on gender differences in the upper tails of test scores.

Perhaps unsurprisingly, given the geographic patterns discussed above, there is a negative correlation between a state’s SAI and its median income level. The coefficient estimate from a simple linear regression of a state’s SAI at the 95th percentile on its median income implies that $10,000 increase in a state's median income decreases the state’s SAI by 0.17, which is significant at the 5% level. That is approximately equivalent to a change of 7 spots on the SAI ranking. We find less of a correlation between gender gaps and the level of adult education in the state. The correlation between the fraction of adults with high-school educations and the SAI is also negative, but is not statistically significant at conventional levels.

We also investigate more direct measures of cultural attitudes and gender stereotypes at the state level using questions from the General Social Survey (GSS). The GSS does not ask questions directly related to gender stereotypes on standardized test scores. However, there are two relevant questions on gender attitudes/issues that have been asked consistently between 1972 and 2006 and provide enough responses in most states to allow for meaningful statistical association. The first question asks “Is it much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family?”11 The second question asks “do you think it should be possible for a pregnant woman to obtain a legal abortion if the woman wants it for any reason?”12 The first question is clearly related to attitudes about gender roles, while the abortion question is an indicator of cultural conservatism.

Because of the relatively small samples used in the GSS, in order to conduct analysis at the state level it is necessary to pool the data from the three decades of GSS surveys. We furthermore limit our analysis to states in which there were at least 100 survey respondents who answered these

---

11 Pooling across years, 41.5% of respondents answered yes to this question.
12 Pooling across years, 40.1% of respondents answered yes to this question.
questions in the pooled data. This procedure leaves 37 states with adequate GSS data. In order to minimize effects of differential sampling in the GSS data, we first regress the individual responses to each of these two questions on the individual-level demographic variables included in the GSS: age, gender, race, self-reported income, and education. We then average the residuals from this regression at the state-level in order to generate state-level measures of cultural attitudes that are independent of the demographic characteristics of the survey respondents. Finally, we standardize these state-level averages by differencing out the mean across states and dividing by the standard deviation across states so that our results can be interpreted using standard-deviation units.

The top two panels of Figure 4 show the correlations between these standardized-residual measures and the state’s SAI at the 95th percentile. There is a very strong correlation between both of these measures and the level of stereotypical gender gaps in test scores. States where people are more likely to answer that it is better if women take care of the home have higher levels of the SAI. Similarly, states where people are more likely to answer that women should not be allowed to have legal abortions also have higher levels of SAI.

Table 1 provides the coefficient estimates for these correlations by showing results from simple linear regressions of these measures on the state-level SAI at both the 95th and 75th percentiles. In Panel A, Column (4), the regression of SAI at the 95th percentile on the normalized residual for the question about women being better suited reveals that a one-standard-deviation increase in this measure yields a predicted increase in the SAI of 0.214 (p-value < 0.01). That is roughly the predicted effect of a $20,000 decrease in a state’s median income level. Looking at the R-squared, this simple measure of gender attitudes accounts for 44% of the variation in the state-level SAI at the 95th percentile cutoff. The question about abortion has a slightly lower R-squared (33%), but has

---

13 The results are very similar and all the qualitative conclusions are the same if we use the raw GSS answers rather than the residuals in the following analysis.
14 Figure 5 shows that these patterns hold at the 75th percentile as well.
of a similar coefficient estimate that is also highly statistically significant. Interestingly, these measures of gender attitudes have much higher R-squared values than a state’s median income (R-squared of 14% in Column 1). Given the smaller gender gaps at the 75\textsuperscript{th} percentile, it is not surprising that the coefficients on these same regressions are somewhat lower at the 75\textsuperscript{th} percentile, but the patterns at the 75\textsuperscript{th} percentile hold and the coefficients remain highly statistically significant.

These questions from the GSS provide information about the cultural attitudes of adults, but it is also interesting to think about the gender-role attitudes of children taking the tests and how those attitudes might correlate with stereotypical test-score performance. We were able to find one piece of evidence on children’s gender attitudes. An earlier wave of the NAEP given in 1990 and 1992 asked students to say how much they agreed or disagreed with the statement “math is for boys”. We gathered the responses to this question for students who took the math test in 1990 and 1992 and found the percent of students who were undecided or agreed with the question in each state. The third panel in Figures 4 and 5 show the correlation between the answers to this question and the state’s SAI at the two cutoff levels. The regression coefficients are found in Column 6 of Table 1. Although the correlation is somewhat weaker than with the GSS questions, there is a significant positive correlation between the fraction of students agreeing with the statement “math is for boys” and the state’s SAI.

\textbf{Robustness.} We are able to provide robustness checks for two potential confounders in our analysis. First, one potential worry is that there is some small underlying variation in the ratio of boys to girls at the state level who actually take the NAEP tests. It is unclear whether this variation is driven by differential rates of public school attendance by gender or by underlying sex-ratio differences at the state level.\textsuperscript{15} In order to address this issue, we replicate the analysis above, netting

\textsuperscript{15} One potential worry for example would be that parents from certain states are more likely to send their girls to private schools than parents from another state. Of course, in order for this to cause the patterns that we find, parents from certain states would have to be more likely to send their girls who have a propensity for math and
out the relevant gender ratio for all test takers in the state. So for example, if a state has a male-female ratio of 1.75 on the math test at the 95th percentile and a male-female ratio of 1.02 overall among math-test takers, the state’s net male-female ratio would be 1.73. The results above hold throughout, and are actually strengthened in many cases, when we account for the underlying variation in gender ratios.

A second potential confound that may exist is the extent to which our analysis is driven by the choice of a national cutoff. While the national distribution is generally the one of interest in discussions of the gender representation in top scientific fields, an alternative approach is to look at the gender ratios in the upper tails of each state’s individual test-score distribution. One reason to investigate state-level cutoffs is that when using the national cutoff, in low-performing states, we are analyzing gender ratios in very high upper tails of the states distribution. In order to address this issue, we replicate the analysis above using a cutoff of the 95th and 75th percentile within each state. The findings above hold throughout when using these state-level cutoffs instead of the national cutoff.

II. Discussion and Conclusion

This paper adds important new information to the literature on gender differences in test scores at the high end of the distribution in the United States. First, we find that although the prevailing national patterns hold throughout the country, there is substantial variation in the gender ratios in the high end across states. This variation tends to be geographically clustered. Second, we show that states with highly unequal ratios in favor of boys on math and science tests also tend to have highly unequal ratios in favor of girls on reading tests. This suggests that the causes of gender science to private school while at the same time be less likely to send their girls who have a propensity for reading to private schools. While this may seem unlikely, we are still able to control for this type of behavior.
inequality in high performance on test scores are more likely to stem from stereotyping and the imposition of gender roles, than from broadly better treatment of one sex over the other. Finally, when we investigate the correlates of adherence to gender stereotypes on tests at the state level, we find that states with lower median income and less educated adult populations have greater gender disparities, and that indicators of cultural gender attitudes are highly correlated with stereotypical gender differences on test scores.

Our analysis has several limitations that point to directions for future research. For example, while we are able to analyze a significant portion of the upper tail of the test score distribution, data limitations make it difficult for us to analyze variation in gender ratios for students scoring in the very highest percentiles (99th percentile or higher). Because this range of extreme talent is relevant for discussions of gender representation in very competitive fields such as scientific academia, it will be important to extend this analysis to higher percentiles of test scores as more data become available. Also, while we argue that our results indicate the importance of environmental factors in contributing to the test score gap, we are unable to identify the exact cultural or environmental differences that may be driving the results we find. Possible candidates include differences in resource allocation, home or classroom instruction, opportunities in the workforce, or simply the psychological effect of stereotypes. Future research can hopefully examine more deeply and develop policies to mitigate the specific forces that influence stereotypical gender disparities in test scores.
References


Figure 1. Male-female ratios in Science, Math, and Reading Across the Distribution

Notes: This figure uses data from the National Assessment of Education Progress. All white, 8th-grade students who took the test between 2000 and 2005 in Science, Math or Reading are included in the sample. Male-female ratios were created at each five percentile level using the sample weights provided in the data.
Figure 2. The Gender Gap in Math and Science and the Gender Gap in Reading by State

Notes: This figure illustrates the relationship between male to female ratios in math and science and female to male ratios in reading by U.S. states. Panel A computes ratios by looking at students scoring in the top 5% while Panel B focuses on students scoring in the top 25%, using national-level cutoffs in each case.
Notes: These maps present the Stereotype Adherence Index (the average of the male-female ratios in math and science and the female-male ratio in reading) for the top 5% (Panel A) and top 25% (Panel B) of students. States are ordered by this Index and then broken into four categories. Each shade of color represents a different grouping with the darker shades indicating a larger amount of stereotypical gender differences.
Figure 4. Attitudes on Gender Issues and Stereotype Index - Top 5%

Notes: This figure shows the relationship between the Stereotype Adherence Index (a measure of the stereotypical gender differences on test scores) and attitudes on women's issues. Panel A, graphs the SAI against normalized, residual responses to a GSS question asking whether women are better suited to stay at home and Panel B for a question about whether abortion is okay. Panel C graphs the SAI against 8th grader responses to "math is for boys".
Figure 5. Attitudes on Gender Issues and Stereotype Index - Top 25%

Notes: This figure shows the relationship between the Stereotype Adherence Index (a measure of the stereotypical gender differences on test scores) and attitudes on women’s issues. Panel A, graphs the SAI against normalized, residual responses to a GSS question asking whether women are better suited to stay at home and Panel B for a question about whether abortion is okay. Panel C graphs the SAI against 8th grader responses to "math is for boys".
Table 1. Correlates with Stereotypical Gender Differences at the State Level

### Panel A

<table>
<thead>
<tr>
<th>2000 Census Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median household income ($1,000s)</td>
<td>-0.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.009)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction with HS degree</td>
<td>-1.822</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.417)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction of females with HS degree</td>
<td>-1.875</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.434)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey Questions</td>
<td>0.184</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abortion not okay for any reason</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(normalized residuals)</td>
<td>(.058)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women better suited for home</td>
<td>0.214</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(normalized residuals)</td>
<td>(.042)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undecided or Agree Math is for Boys</td>
<td>0.107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(normalized residuals)</td>
<td>(.039)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| R-squared                             | 0.144 | 0.062 | 0.065 | 0.325 | 0.440 | 0.111 |
| Observations                          | 37    | 37    | 37    | 37    | 37    | 35    |

### Panel B

<table>
<thead>
<tr>
<th>2000 Census Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median household income ($1,000s)</td>
<td>-0.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.003)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction with HS degree</td>
<td>-0.988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.454)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction of females with HS degree</td>
<td>-0.966</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.462)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey Questions</td>
<td>0.069</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abortion not okay for any reason</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(normalized residuals)</td>
<td>(.013)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women better suited for home</td>
<td>0.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(normalized residuals)</td>
<td>(.012)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undecided or Agree Math is for Boys</td>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(normalized residuals)</td>
<td>(.016)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| R-squared                             | .426 | .159 | .149 | .400 | .331 | .205 |
| Observations                          | 37    | 37    | 37    | 37    | 37    | 35    |

**Notes:** This table illustrates the relationship between the Stereotype Adherence Index (a measure of the stereotypical gender differences on test scores) based on the top 5% (Panel A) and the top 10% (Panel B) and state characteristics including attitudes on women's issues. The abortion and women better suited for home questions are normalized residuals taken from the General Social Survey and the "math is for boys" question is taken from an earlier NAEP wave.