
Volume 42, Issue 2, March 2013
ISSN 0049-089X
ELSEVIER
Social Science
RESEARCH
A Journal of Social Science Methodology and Quantitative Research
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Social Science Research 42 (2013) 432–445

Contents lists available at SciVerse ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

Educational standardization and gender differences in mathematics achievement: A comparative study

Hanna Ayalon [a,*], Idit Livneh [b]

[a] Department of Sociology and Anthropology, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel
[b] School of Education, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel

ARTICLE INFO

Article history:
Received 2 August 2011
Revised 10 October 2012
Accepted 10 October 2012
Available online 23 October 2012

Keywords:
National examinations
Between-teacher instructional variation
Gender stratification
Economic development
TIMSS
Hierarchical analysis

ABSTRACT

We argue that between-country variations in the gender gap in mathematics are related to the level of educational system standardization. In countries with standardized educational systems both genders are exposed to similar knowledge and are motivated to invest in studying mathematics, which leads to similar achievements. We hypothesize that national examinations and between-teacher uniformity in covering major mathematics topics are associated with a smaller gender gap in a country. Based on Trends of International Mathematical and Science Study (TIMSS) 2003, we use multilevel regression models to compare the link of these two factors to the gender gap in 32 countries, controlling for various country characteristics. The use of national examinations and less between-teacher instructional variation prove major factors in reducing the advantage of boys over girls in mathematics scores and in the odds of excelling. Factors representing gender stratification, often analyzed in comparative gender-gap research in mathematics, are at most marginal in respect of the gap.

1. Introduction

The gender gap in mathematics achievement has been the subject of extensive research. For long, explanations for boys' advantage over girls referred to individual factors, such as biological influences, attitudes to mathematics and science, and familial influences (Davenport et al., 1998; Xie and Shauman, 2003), and to school factors, such as course-taking and encouragement by teachers and counselors (Xie and Shauman, 2003). A significant rise in the availability of comparative data has enriched this research by adding structural features of countries as well as norms and values as factors that may explain variations in this gap. Most comparative gender-gap research in mathematics achievement focuses on between-country differences in women's participation in various spheres of life, such as education, the labor market and politics (e.g., Else-Quest et al., 2010; Fryer and Levitt, 2010; Penner, 2008) and on economic development (e.g., Guiso et al., 2008; Riegle-Crumb, 2005). Surprisingly, this research hardly refers to traits of national educational systems.

We attempt to fill this void by introducing the level of educational standardization as a country characteristic that may be linked to the gender gap in mathematics achievement. We base this line of research on previous findings which reveal a negative association between the level of educational standardization and socioeconomic inequality in math achievement. We proceed as follows: First, we review of the literature that studies the relations between gender stratification, economic development, and the gender gap in mathematics achievement. Second, we develop our hypothesis on the link between educational standardization and the gender gap. Third, we describe the data, the analysis, and the findings. We conclude by discussing some implications of the study.

* Corresponding author. Address: Dep. of Sociology, Tel Aviv University, Tel Aviv 69978, Israel. Fax: +972 36409215 (office), +972 36344941 (home).
E-mail address: ayalon@post.tau.ac.il (H. Ayalon).

0049-089X/$ - see front matter
http://dx.doi.org/10.1016/j.ssresearch.2012.10.001

2. The gender gap in mathematics, gender stratification and economic development

Comparative research on the gender gap in mathematics is often based on the gender stratification hypothesis (Else-Quest et al., 2010), suggesting that structural and cultural factors play a role in the formation of this gap. Countries differ in degree of economic development and in level of women's participation in various spheres. This will affect girls' exposure to mathematics, hence the gender gap in achievement. Researchers have studied a variety of effects on the gap: women's rates of participation in education, the labor market, and politics (Baker and Perkins Jones, 1993; Fryer and Levitt, 2010; Hyde and Mertz, 2009; Penner, 2008; Riegle-Crumb, 2005); the primacy for women of children and home life (Penner, 2008; Riegle-Crumb, 2005); equality in labor-force position (Penner, 2008); women's share in research jobs (Else-Quest et al., 2010) and economic development (Guiso et al., 2008; Riegle-Crumb, 2005). This line of research is based on the assumption that lower rates of participation of women in education, in the labor market and in politics, primacy of home and family in women's lives, as well as disadvantaged positions in the labor market, produce gender stratification by according women lower status and dependent positions. Gender stratification influences the behavior of youth and of parents and educators involved in their socialization and academic training (Baker and Perkins Jones, 1993; Else-Quest et al., 2010; Riegle-Crumb, 2005). In gender-stratified countries female students incline by their own choice or by their schools' encouragement to follow the traditional pattern of perceiving mathematics as less important for their future, so they put less effort into mathematics courses or avoid them altogether (Baker and Perkins Jones, 1993).

The findings generally confirm the assumption that gender equality in a country correlates with the gender gap in mathematics, but some results are contradictory and sometimes counterintuitive. Higher levels of women's participation in politics are indeed accompanied by a narrowed gender gap in mathematics (Guiso et al., 2008; Else-Quest et al., 2010; Penner, 2008; Riegle-Crumb, 2005). Else-Quest and colleagues found the expected positive effect of gender equity in school enrollment on gender equality in mathematics achievement. It is different for women's participation in the labor market. Riegle-Crumb did not find a significant effect of this factor on the gender gap, whereas Penner found that women's increasing labor market participation was accompanied by a widening rather than a narrowing of the gap. Penner explains this counterintuitive result by the link between women's participation in the labor market and gender segregation. Countries with higher percentages of women in the labor market evince higher levels of labor-market gender segregation (e.g., Mandel and Semyonov, 2005), which sustains the traditional image of women as less interested in and less qualified for mathematics and sciences. Contrary to this, Baker and Perkins Jones, and Guiso and coworkers report that an increase in women's economic activity is related, as expected, to the moderation of the gender gap.[1] Else-Quest and collegues (2010) refer to a particular angle of labor market participation – women's share in research jobs – and report that their higher share moderates boys' advantage in math achievement. Another aspect of gender stratification, the primacy of family and home life for women, was included in Riegle-Crumb's and in Penner's analyses. Both studies, which used different indicators,[2] found no link between this feature and the gender gap in math.

A different approach was adopted by Guiso and colleagues (2008) and by Fryer and Levitt (2010). In their analyses these researchers incorporated a comprehensive composite index of gender equity.[3] Both studies used the same indicator, but each reached different results. Guiso and colleagues found the expected negative effect of the index on the advantage of boys over girls, whereas Fryer and Levitt found no effect of the index on the gender gap. Fryer and Levitt ascribed the different results to the different countries covered by the two analyses, in particular that Middle Eastern Muslim countries, which were included in their sample, were absent from Guiso et al.'s. This difference is very important, as we will elaborate later.

Another result that did not follow the predictions is the effect of countries' economic development on the gender gap in mathematics. By modernization approaches, a positive link is expected between economic development and gender equality in mathematics and science achievement. Economic development is expected to be linked to egalitarian values, which may cause female students to perceive mathematics as relevant to their future careers. However, Riegle-Crumb (2005) found no effect of a country's GDP per capita on the gender gap in mathematics. This also holds for most analyses performed by Guiso et al. (2008). This result partially accords with the findings of Charles and Bradley (2009), who, in studying sex segregation in higher education found a positive rather than the expected negative effect of GDP on it. These authors explained this counterintuitive result as the consequence of two cultural forces which produce "a new sort of sex segregation regime" in developed societies: a gender-essentialist ideology and self-expressive value systems. They argue that the centrality of gender in establishing identity creates a self-expression value system that encourages the development of culturally masculine or feminine affinities. This may make girls averse to mathematics and avoid related programs where self-expression is a legitimate, even normative, criterion for curricular choice.

In sum, despite the value of gender stratification and economic development for comparative research, these country characteristics offer limited and sometimes contradictory explanations of between-country variations in the gender gap in mathematics achievement.

1 Baker and Perkins Jones (1993) present bivariate correlations, whereas Guiso et al. (2008) control only for GDP per capita. This suggests that the moderating effect of women's economic activity may be spurious, reflecting other components of gender stratification.

2 Riegle-Crumb (2005) used fertility rate and availability of legal abortion; Penner (2008), who defined the aspect as domesticity, used variables such as the amount of time women spend on housework and the importance of children for women's fulfillment.

3 The index (World Economic Forum's Global Gender Index: WEF-GGI) is composed of measures of economic participation and opportunity, educational attainment, political empowerment, health and survival.

3. The gender gap in mathematics and the level of educational standardization

Focusing on between-country variations in gender equality, comparative research on the gender gap in mathematics has rarely referred to features of national education systems, although their relevance to educational outcomes is self-evident. Countries vary in how their educational systems are organized, and this affects the reality that students experience at school. The structural features of an educational system may be linked to gender parity no less—and perhaps even more—than the macro-level indicators of gender stratification, which are usually used in this research.

Contrary to the marginality of structural features of educational systems in comparative research on gender inequality in mathematics achievement, they are central in between-country comparisons of socioeconomic inequality. A structural feature which appears significantly linked to variations in socioeconomic inequality in achievement is the level of standardization, "the degree to which the quality of education meets the same standards nationwide" (Allmendinger, 1989, 233). Standardization is usually defined by the use of a national curriculum and national examinations, and standardized school resources (Allmendinger, 1989; Park, 2008; Van de Werfhorst and Mijs, 2010). Van de Werfhorst and Mijs, who reviewed the comparative literature on the impact of national educational institutions on socioeconomic inequality in achievement, concluded that two components of standardization, national curriculum and especially the use central examinations, decrease inequality.

The equalizing effect of a national curriculum and national examinations on socioeconomic inequality has two major sources: transparency of the educational structure and its rules, and lesser differentiation in exposure to knowledge.

When educational decisions are based on scores in national standardized tests this affects students, parents and teachers. Students and parents of different social strata understand the value of succeeding in these tests, and aim for high scores. The lower ability of less educated parents to decode educational structures and their rules is less relevant in standardized systems (Park, 2008). The clear standards and information prevalent in standardized systems prevent teachers from basing educational decisions such as course placement on subjective assessment (Muller and Schiller, 2000). Fuchs and Woessmann (2007) underscore the value of national examinations by showing that even when a school is autonomous in its curriculum, examinations direct teachers in what to teach. Ayalon and Gamoran (2000) showed that standardization may interact with differentiation. In standardized systems, the national examinations motivate lower-track students to achieve better scores, thus eliminating the usual stratifying effect of differentiation. Stevenson and Baker (1991), in their comparative study based on the Second International Mathematics Study (SIMS), concluded that in the absence of state control of the curriculum (i.e., control of curriculum contents and of examinations) teachers tend to modify it according to their needs or their students', thereby forging a link between students' characteristics and their exposure to knowledge. The latter is a major determinant of achievement: students who learn more get better scores. This is true for many school subjects, but particularly for mathematics (e.g., Gamoran and Hannigan, 2000; Riegle-Crumb, 2006). In fact, course-taking is the best predictor of mathematics achievement (McFarland and Rodan, 2009).

Montt (2011), who defines curriculum standardization as a component of the students' opportunities to learn, reports on different results. Using PISA 2006, he did not find a significant effect of this factor on achievement inequality across nations, defined as the variance of student math scores. Montt assigns these results to low levels of variance in curriculum standardization, measured by the degree of the central government's control of the curriculum. His results may also stem from the dependent variable, which, unlike most research in this topic, measures the variance of achievement.

National standardized examinations are widely discussed in the US in the framework of the debate over high-stakes testing and accountability policies. The research on this topic yields contradictory results. Several between-state comparative studies show that students in high accountability states have better achievement gains in math tests (Carnoy and Loeb, 2002), and that the effect of school's accountability depends on the state's degree of local control (Loeb and Strunk, 2007). Amerin and Berliner (2002) claim, however, that due to insufficient data, straightforward conclusions on the effects of high-stakes testing are premature. Qualitative studies, mainly concentrating on a single community (e.g., Diamond, 2007; Watanabe, 2008) suggest that high stakes testing has a positive effect on the instruction of high performing schools only. Still, even critics of high stakes testing agree that this policy enhances uniformity in instructional content.

We argue that the mechanisms that produce the equalizing effect of standardization on socioeconomic inequality in achievement also work for gender inequality. A common explanation for the gender gap in mathematics achievement is girls' perception that the subject is irrelevant for their future career (Baker and Perkins Jones, 1993; Else-Quest et al., 2010; Xie and Shauman, 2003). This reasoning, which is expected to be particularly true in gender-stratified countries (Riegle-Crumb, 2005), seems less persuasive for educational systems where test scores, based on national standards, have clear implications for immediate educational outcomes, such as placement in secondary education and eligibility for a diploma. Even girls who view mathematics as irrelevant for their future occupational career will be motivated to do well in this subject if their achievements have immediate and clear consequences for their progress in the educational system. The same holds for parents and teachers. Even those who perceive mathematics as less relevant for girls will be less likely to encourage them to jeopardize their future by neglecting it. This implies that in standardized educational systems girls will have better motivation to succeed in this subject. This line of reasoning wins some support from Tsui (2007), one of the few researchers who attributes variations in the gender gap in mathematics achievement to structural features of educational systems. Comparing China and the US, Tsui reports gender parity in average achievement among high school graduates in the former, and outperformance by boys in the latter. In explaining this pattern, Tsui refers to China's rigorous national mathematics curriculum, which imposes a universal standard and includes national examinations, and the absence of these institutions in the US.[4]

Standardized systems increase girls' exposure to mathematics. In his analysis of students' achievements in SIMS, Garden (1989) attributes girls' lower achievement to the choice by disproportionately high numbers of mathematically able girls not to pursue this subject. The value of exposure to mathematics is clearly demonstrated in within-country research. Xie and Shauman (2003) report that course-taking is a central predictor of girls' achievement in mathematics. Hyde and Mertz (2009), who analyzed changes in the patterns of mathematics course-taking in the US, show that increasing gender equality in course-taking is linked to a reduction, even the elimination, of the gender gap in achievement.

Following this logic, our study posits that standardized educational systems, which practice a national curriculum, hold national examinations and expose students to similar math knowledge, will be linked to the moderation or even elimination of the gender gap in mathematics achievement, regardless of the degree of gender stratification and economic development.

4. Data and method

4.1. Data

We use the Trends of International Mathematical and Science Study (TIMSS), conducted in 2003. This is an international assessment project run by the International Association for the Evaluation of International Achievement (IEA). We preferred TIMSS to the Programme for International Student Assessment (PISA), another well known project of international assessment, because TIMSS tests knowledge in mathematics according to curriculum, not general mathematical literacy, which is tested by PISA. TIMSS is also based on a more diverse sample, including developing countries.

The target population in 2003 was fourth- and eighth-graders. The sample design for TIMSS is a two-stage stratified cluster. The first stage is taking a sample of schools; the second is taking one or more classes out of the target grade in the sampled schools. Weights are used to obtain nationally representative results. Our data on mathematics achievements, student's gender, and characteristics of the educational systems are based on TIMSS.[5] Variables representing economic development and gender stratification are taken from UNDP (United Nations Development Data),[6] which includes development indicators for a variety of countries. We mostly use data for 2004.

4.2. Sample

We restricted the analysis to students who took the exam in the eighth grade because the sample of countries whose fourth graders were tested and yielded enough data for the current analysis was very small: 15 countries in all. For our analysis we chose 32 of the 46 countries that provided data for eighth graders. The choice was based on substantive and practical considerations, namely the presence of a variety of cultural contexts, which is essential for comparative research in general (Else-Quest et al., 2010; Park, 2008). In our case, we believe that analysis of countries which differ in the degree of gender stratification is essential for assessing whether curricular standardization operates similarly in different gender regimes. Differently stated, the heterogeneity of the sample will help in assessing whether standardization moderates gender inequality regardless of the country's values regarding gender equality. Countries whose data were not in UNPD and countries whose data on the educational variables in TIMSS were incomplete had to be excluded. The sample included developed and developing countries. The final sample was 137,666 students distributed among the 32 countries. The list of countries and some of their characteristics are presented in Table 1.

4.3. Variables

4.3.1. Student variables

4.3.1.1. Achievement. We analyzed the gender gap in mean mathematics scores and in the odds that a student was in the top 10% and the bottom 10% in his/her country (see also Penner, 2008; Xie and Shauman, 2003). Test scores were standardized to an international average of 500 and standard deviation of 100.

4.3.1.2. Explanatory variables. The student-level explanatory variable was gender, coded 1 for males.[7]

These variables have no missing cases.

[4] But Tsui (2007) found that gender inequality in excelling in math was common to the two countries.

[5] http://timss.bc.edu/timss2003.html.

[6] http://hdr.undp.org/en/statistics/data.

[7] The data include a variety of student variables, such as age, family size, immigrant status, parental education, number of school-related possessions at home, number of books at home. We considered their inclusion in the analysis as controls, but since they do not correlate with gender in either country, and our major interest is the gender slope, their inclusion has no substantive contribution to the analysis. To test this point, we performed additional analyses with these variables as additional student characteristics. Since the results of the restricted and the extended models were very similar, we are presenting only the restricted, parsimonious, model.

Table 1
Selected characteristics of the countries.

	N	Mean scores	Difference Boys–Girls	Odds ratio (boys/girls) top 10%	Odds ratio (boys/girls) bottom 10%	Use of national exams	Between-teacher instructionalvariation[a]	% seats in parliament held by women	Women's share in the labor force
Bahrain	4071	405.79	−31.37*	.78*	4.26*	Yes	.21	13.8	18.2
Belgium	4956	544.18	9.17*	1.33*	1.06	No	−.35	35.7	43.1
Botswana	4096	375.59	3.03	1.16	1.03	No	−.12	11.1	42.6
Bulgaria	4065	484.52	−6.05*	1.21	1.30*	No	−.55	22.1	46.5
Chile	6169	405.87	14.55*	1.42*	.77*	Yes	−1.20	12.7	34.4
Egypt	6819	442.68	7.39*	1.50*	1.05	Yes	.75	3.8	22.7
Estonia	3878	535.38	−5.07*	.82	1.14	Yes	−1.61	21.8	49.4
Ghana	3585	293.84	22.99*	1.78*	.71*	No	1.21	10.9	48.1
Holland	3033	540.54	6.20*	1.49*	.80	No	.27	36.0	44.1
Indonesia	5653	420.11	1.69	1.13	1.11	Yes	.16	11.3	37.7
Iran	4685	417.19	−5.95*	1.01	1.50*	No	.24	4.1	31.6
Israel	4301	494.48	4.57*	1.74*	1.22*	Yes	1.20	14.2	46.2
Italy	4270	482.35	6.83*	1.67*	1.05	Yes	.11	16.1	39.0
Japan	4809	568.50	2.97	1.41*	1.34*	Yes	−.01	11.1	40.9
Jordan	4480	423.34	−26.59*	.61*	2.33*	Yes	.11	7.9	24.1
Korea	5307	587.61	5.06*	1.23*	1.11	Yes	.19	13.4	40.0
Lebanon	3441	444.17	13.64*	1.67*	.63*	No	−1.01	4.7	29.3
Latvia	3567	518.14	−2.45	1.31*	1.68*	No	.73	19.0	48.8
Lithuania	4519	507.55	−2.30	1.26*	1.38*	Yes	−1.16	24.8	49.4
Macedonia	3886	436.86	−8.40*	1.05	1.28*	Yes	.84	28.3	40.1
Malaysia	5314	507.87	−8.15*	1.04	1.54*	Yes	.21	13.1	35.2
Moldova	4000	462.41	−7.64*	.82	1.46*	Yes	.42	21.8	47.6
Morocco	2402	393.60	10.53*	1.50*	.65*	No	2.60	6.4	25.9
New Zealand	3799	487.82	−5.97*	1.05	1.35*	No	−.39	32.2	45.8
Norway	4133	461.04	−.85	1.10	1.15	Yes	−2.40	37.9	47.2
Philippines	6270	383.84	−10.03*	1.04	1.26*	Yes	.06	22.1	38.7
Romania	4021	481.18	−3.94	1.25*	1.07	No	−.36	10.7	45.2
South Africa	6175	288.69	16.58*	1.49*	.90	No	.85	32.8	38.6
Slovenia	573	482.34	−12.24*	.63	1.09	No	−1.32	10.8	45.9
Sweden	4250	498.59	1.92	1.22*	.94	Yes	−.50	47.3	47.7
Tunisia	4315	414.39	23.72*	1.60*	.27*	No	1.65	19.3	26.7
United Kingdom	2,824	506.75	1.90	1.05	1.10	No	−0.82	19.3	45.4
Total	137,666	455.96	3.35*	1.21*	1.13*		.00	18.6	39.56

* $p < .05$.
[a] Composed of coefficients of variation of teachers' answers on time they devote to 5 math topics. See explanation in text and details in Table A1 (in the Appendix).

4.3.2. Country variables

4.3.2.1. Measures of standardization. As noted, standardization is usually measured as use of a national curriculum and of national examinations (Van de Werfhorst and Mijs, 2010). Since all countries in our sample except Belgium reported having a national curriculum, we measured standardization as use of national examinations. The information on the use of this practice was provided by the national research coordinators, who worked with experts of the curriculum in each country, such as curriculum supervisors of mathematics. The specific question asked whether national assessment was used to help the implementation of the national mathematics curriculum. National examinations contribute to uniformity in teaching, but there still may be variations in the way teachers implement the curriculum in their classes (Resh and Benavot, 2009). To cover this aspect of standardization we constructed an index of between-teacher instructional variation in the time devoted to major mathematics topics. Teachers were asked how often they referred to five mathematics topics (addition, fractions, problems, data tables, equations). The answers ranged from never, through some lessons and about half of the lessons, to almost all or all lessons. We computed the coefficient of variation for each topic. Principal component analysis showed that all items were highly loaded on one component. Cronbach's alpha for reliability of the items was .73. We constructed a composite variable of the five coefficients of variation, weighted by the principal component scores. We named the variable between-teacher instructional variation.

4.3.2.2. Economic and human development (development). A composite variable of Human Development Index (HDI), which is a combination of indicators of life expectancy, educational attainment and standard of living, calculated by the Human Development Report Office (HDRO).[8] To broaden the index we added percent of urban population to the existing composite variable.

4.3.2.3. Gender stratification. Following Riegle-Crumb (2005), gender stratification was represented by three domains: the labor market, politics and the family. These aspects were also analyzed by Guiso et al. (2008) and by Penner (2008).[9] As noted, Fryer and Levitt (2010) and Guiso et al. used a composite variable of women's economic and political opportunities, education and well-being. Because previous research (Else-Quest et al., 2010; Penner, 2008) shows that specific domains of gender equity have different effects on the gender gap in math we preferred to analyze the various aspects separately. Women's economic participation is a composite variable of women's economic activity (percent women aged 15 and older in the labor market) and women's share in the labor force. Women's political empowerment is a composite variable of percent women in parliament and in government.[10] Primacy of the family (family) is a composite variable of total fertility (used also by Riegle-Crumb) and adolescent fertility rate. We added the latter rate to the composite variables because we assumed that giving birth at an early age has special significance for the primacy of children and family.[11]

The variables and the descriptive statistics are presented in Table A1 in the Appendix.

4.4. Method

Since we analyzed students nested within countries we conducted hierarchical analysis using HLM 6.08. We performed linear analysis for mean mathematics scores and binary logistic for the odds of belonging to the two extremes of the distributions. We weighted the student variables with student weights (SENATE) provided by the data.[12] HLM 6.08 uses a method of computation devised by Pfeffermann et al. (1998) for hierarchical data. This method is based on weighting the information of each case in the framework of maximum likelihood.

We centered all student and country quantitative variables around their grand means. The dummy variables retained their original form. We allowed the slope of gender, which represents gender inequality in mathematics achievement, to vary among countries.

The models we estimated are

$$outcome_{ij} = \beta_{0j} + \beta_{1j}(gender) + e_{i_j}$$

where β_{0j} is the average outcome in country j, β_{1j} is the effect of gender in country j, e_{ij} is the error term for student i in country j (relevant only for scores).

The equation for the gender slope is

8 http://hdr.undp.org/en/media/HDR_2011_EN_Tables.pdf.

9 Penner (2008) also analyzed education and labor force equality. We could not include these two aspects in the analysis because they are highly correlated with the development index. Penner did not face this problem because he did not include a measure of development in his analysis.

10 Similar measures are used by Riegle-Crumb (2005), Penner (2008), and Guiso et al. (2008).

11 Since we analyzed only 32 countries, we were very cautious regarding the number of country variables, particularly the significantly intercorrelated. Thus we omitted from the analysis several relevant variables because they were highly correlated with development: women's enrollment in education (correlation of .81), the gender empowerment measure (.77), and the ratio of female to male income (.61). We considered the inclusion of the last variable in the index of economic participation, but in view of Penner's finding that high rates of female participation in the labor market are related to an increase in girls' disadvantage in math, we resolved that labor market participation and income inequality constitute two different domains regarding the gender gap in math achievement, hence should not be incorporated in the same index.

12 The data provide additional weights. We used SENATE, which is recommended for comparative analysis. Our findings were not sensitive, however, to other choices of weights.

$$\beta_{1j} = \gamma_{10} + \sum_{1}^{l} \gamma_{1j} Z_j + u_{1j}$$

where γ_{10} is the average link between gender and outcome, $\sum_1^l \gamma_{1l}$ is the sum of the coefficients of L country variables (non-educational variables in the first equation; educational structure variables added in the second equation), u_{1j} is the error term for country j.

Each analysis covered two models. In the first model, gender stratification and human and economic development were included in the equations of the intercept and of the gender slope. The variables representing educational structure were added to the equation in the second model. Since top10% and bottom 10% refer to each country, between-country variation of the intercept of the equations of these outcomes does not have substantive meaning. We included the country variables in the equations of the intercepts of these outcomes because in multilevel models the variables included in the equation of the slopes have to be included in the equation of the intercept (Raudenbush and Bryk, 2002); but we do not present or discuss these models which, as noted, lack substantive meaning.

5. Results

5.1. The countries and their characteristics

For a better sense of the countries' positions regarding the gender gap, Table 1 presents their average scores, the three measures of inequality in achievement (difference in boys' and girls' scores, and odds of boys versus girls belonging to the top 10% and the bottom 10%), and four of the explanatory variables: use of national exams; between-teacher instructional variation, gender stratification, represented by percent seats in parliament held by women and by women's share in the labor force.[13]

The table shows the variation in gender inequality in all three outcomes, and the differences in the three outcomes. Boys have higher scores than girls in most countries, but the differences are modest, as already reported by previous research (Else-Quest et al. ,2010; Guiso et al., 2008). The largest difference between the two genders is in favor of girls (in Bahrain, were it reaches about one third of the standard deviation). Boys' advantage reaches at most about one quarter of the standard deviation (in Tunisia and Ghana). Girls' disadvantage is clearer in the higher part of the distribution: the odds ratio exceeds 1 (indicating an advantage of boys over girls) in most countries; in 19 countries it reaches statistical significance. Only in Bahrain and Jordan are the odds ratios significantly smaller than 1, indicating an advantage of girls over boys. Boys have no advantage over girls at the bottom of the distribution: on average, boys' odds of failure are 1.13 greater than girls'; the odds are not significantly different from 1 in 14 countries, and boys have higher odds of failure in 12 countries. Only in six countries are the odds of failure significantly higher for girls.

Gender inequality in math achievement does not necessarily follow the country's tradition regarding women's status. Muslim countries are conservative genderwise, as manifested in the table by the low percentages of women in parliament and in the labor market. Boys score significantly higher than girls in four Muslim countries: Tunisia, Morocco, Lebanon and Egypt, but girls have higher scores in three countries: Iran, Jordan and Bahrain. As noted, only in Jordan and Bahrain do the girls have significantly higher odds of excelling in math. These countries are patriarchal and non-egalitarian regarding gender (Al Mahadin, 2004; Fryer and Levitt, 2010; Mehran, 2003[14]). The parliaments of these countries and their labor markets are notably "male" (women constitute 7.9% of Jordan's parliament, 4.1% of Iran's and 13.8% of Bahrain's; the respective values for share in the labor force are 24.1%, 31.6%, and 18.2%). The advantage of boys over girls is the clearest in the Western countries, which are usually perceived as more egalitarian in terms of gender, and as the table shows, the proportions of women in their parliaments and in the labor market are relatively high. This descriptive information suggests that the link between gender stratification and the gender gap in math achievement, which was mainly reported in research on Western countries, does not hold for more heterogeneous contexts. This proposition gets support from Fryer and Levitt (2010), who report that the link between females' opportunities in a country and the gender gap in math, which they found in Western countries, disappeared when Middle Eastern countries were added to the sample (see also Else-Quest et al., 2010). We will elaborate on this paradoxical pattern in discussing the results of the multivariate analysis.

5.2. Bivariate correlations

Student and country correlation matrices appear in Table A2 in the Appendix. Most correlations are in the expected direction. The correlation between national examinations and between-teacher instructional variation is negative, as expected, but small and statistically insignificant, suggesting that they represent different aspects of standardization. The degree of between-teacher instructional variation in the time devoted to various math subjects does not necessarily follow official

[13] In the analysis we represent women's political empowerment by an index composed of percentage of seats in parliament and in government held by women, and women's economic participation by an index of economic activity and share in the labor market. We prefer to present the original variables because their values make more sense. We present only one the variables representing each aspect because the components of the composite variables are highly correlated (.82 and .85, respectively).

[14] The educational advantage of girls over boys in Jordan is well known, and documented in previous research (e.g., Al-Nhara, 1999).

Table 2
Results of the hierarchical analyses.

	Scores		Top 10%[a]		Bottom 10%[a]	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Student variables						
Intercept	449.958* (5.906)	449.587* (5.906)	−2.384* (.041)	−2.433* (.043)	−2.180* (.045)	−2.086* (.052)
Gender – male	.064 (2.075)	3.295** (1.903)	.178* (.049)	.231* (.042)	.102 (.075)	−.036 (.086)
Country variables						
Effects on the intercept						
Economic and human development	26.373* (5.838)	24.446* (5.833)				
Women's economic participation	26.086* (7.947)*	24.972* (8.581)				
Women's political empowerment	−20.122* (6.837)	−19.794* (6.866)				
Primacy of family	−44.091* (7.680)	−43.907* (7.160)				
National examinations		13.155 (9.093)				
Between-teacher instructional variation		−2.669 (6.806)				
Effects on the gender slope						
Economic and human development	−2.369 (3.441)	.186 (2.433)	−.007 (.080)	.062 (.060)	.163 (.126)	.082 (.091)
Women's economic participation	1.273 (3.563)	3.855 (2.964)	.016 (.077)	.092 (.068)	.035 (.135)	−.039 (.110)
Women's political empowerment	.983 (1.994)	.882 (1.511)	−.002 (.053)	−.002 (.032)	−.098 (.066)	−.093** (.054)
Primacy of family	−.381 (3.125)	−.568 (2.646)	−.005 (.075)	−.009 (.061)	.084 (.106)	.090 (.092)
National examinations		−7.149* (2.998)		−.120 (.074)		.296* (.115)
Between-teacher instructional variation		5.561* (2.007)		.162* (.047)		−.164* (.072)
Statistics for the models[a]						
Intraclass correlation:						
Scores: .409						

Statistics for the model of Scores	Variance Component	Chi square	d.f.	p	Proportion of variance explained by the model	Δ (Model 2–Model 1)
Intercept						
Model 1	1115.71	12860.33	27	.00	.75	
Model 2	1137.95	11964.26	25	00	.75	.00
Gender slope						
Model 1	150.11	804.67	27	.00	.00	
Model 2	114.29	624.30	25	.00	.16	.16
Statistics for the model of Top10						
Gender slope						
Model 1	.072	197.28	27	.00	.00	
Model 2	.049	133.95	25	.00	.24	.24
Statistics for the model of Bottom10						
Gender slope						
Model 1	.192	492.29	27	.00	.00	
Model 2	.150	404.25	25	.00	.16	.16

* $p < .05$.
** $p < .10$.
[a] We refer to the intercept of the equation of math scores only. Country variables are included in the equations of the intercepts of top 10% and bottom 10%, but we do not present their effects on the intercepts or the relevant statistics because they have no substantive meaning. See explanation in text.

policy. We wish to refer more broadly to the impressive correlations of between-teacher instructional variation with human and economic development (−.427) and with women's participation in the economy (−.489) and in politics (−.377). Economically and in terms of gender developed countries show closer uniformity among teachers in time devoted to different math topics. Attempting to understand these intriguing negative links, we speculate that in developed countries it may be due to better teacher training, a longer tradition of teaching various math subjects or a more homogeneous student body. Empirical tests eliminate the first two options[15] but provide some support for the last. The data show substantial positive correlations (.646) between the variance of parental education in a country (defined as average of mother and father's years of schooling), represented by the coefficient of variation, and between-teacher instructional variation. After controlling for this variable, the partial correlations of development, women's economic participation and women's political empowerment with between-teacher instructional variation fall to −.068, −.091 and −.108, respectively. All correlations lose their statistical

[15] The data include four measures of teacher's training. All measures are positively correlated with the development index, but only the correlation with "teachers are required to have first degree" reaches statistical significance. This variable is negatively and significantly correlated with between-teacher instructional variation (−.376). Still, controlling this variable does not produce any significant change in the correlation between development and between-teacher instructional variation. Years since establishment of the curriculum, an additional variable included in the data, is not correlated with development, and its correlation with between-teacher instructional variation is positive, but small and statistically insignificant.

significance. The student body in developed countries is more homogeneous socially. This homogeneity, which probably has to do with less variation in scholastic ability, is linked to increasing uniformity in teaching. With a modicum of speculation, we may suggest that uniformity in teaching is linked more to student-body characteristics than to educational policy.

5.3. Hierarchical models

Table 2 presents the results of the hierarchical analyses. For each of the three outcomes—mathematics scores, top 10% and bottom 10%, two models were estimated. The first contains four non-educational country variables: human and economic development, women's economic participation, women's political empowerment and primacy of family. The two measures of standardization are added to the equation in the second model.

Since the focus of the study is between-country gender inequality, our interest is in the variation of the gender slope. For all outcomes the between-country variation of the gender slope is statistically significant. The significance of the standardization variables in explaining this variance is clear. Prior to their inclusion in the equation, country characteristics make no contribution to the reduction of the variance of the gender slope. The standardization variables reduce the variance of the slope in the equations of scores and of bottom 10% by about 16% and in the equation of top 10% by 24%.

The gender slope reveals the advantage of boys over girls in scores and in the odds to excel, although in the equations of scores the coefficient reaches the threshold of statistical significance only after the inclusion of the standardization variables.[16] It is otherwise for the bottom 10%: the gender slope is statistically insignificant in both equations of this outcome. In a country which does not use national examinations and which is average in all other characteristics, boys' advantage over girls in math scores is about 3.3 points. In a country with such characteristics boys' odds of reaching the top 10% are about 1.3 times ($e^{.231} = 1.260$) greater than girls', while boys' odds of failure in math do not differ from girls'.

The two measures of standardization are significantly linked to the reduction of gender inequality. Increase in between-teacher instructional variation is significantly linked to widening of the gender gap in all outcomes. National examinations are associated with narrowing the gap in scores and with increasing boys' disadvantage in the odds of failure. The coefficient of national examinations in the equation of the odds of excelling does not reach statistical significance.

Statistical significance notwithstanding, the size of the coefficients of the measures of standardization may look unimpressive. For example, the coefficient of national examinations in the equation of scores, −7.15, indicates that this practice moderated the gender slope by about .07 of the standard deviation (106); the coefficient of between-teacher instructional variation in the same equation indicates that each score increases the slope by about .05 of a standard deviation. We should bear in mind, however, that the difference between boys' and girls' scores is small to begin with. The gender slope (3.295) is about .03 of the standard deviation, and as we saw in Table 1 the widest difference between the two genders is about one third of a standard deviation (in Bahrain; in favor of girls).

In an average country which does not use national examinations and is one standard deviation above the mean in between-teacher instructional variation the gender gap increases to about 9 points ($3.295 + 5.561 = 8.856$), whereas it diminishes to about 2.7 points, in favor of girls, in a similar country whose between-teacher instructional variation is one standard deviation below the mean ($3.295 - 5.561 = -2.266$). The parallel values for the odds of excelling are 1.48 ($e^{(.231+.162)} = 1.481$), and 1.07 ($e^{(.231-.162)} = 1.071$), respectively, and for the odds of failing .82 ($e^{(-.036-.162)} = .820$ and 1.14 ($e^{(-.036+.164)} = 1.137$).

In an average country that uses national examinations, the boys' advantage in scores disappears completely. In fact, the gender coefficient changes its sign ($3.295 - 7.149 = -3.854$), but it does not reach statistical significance[17]; boys' odds of reaching the bottom 10% are now 1.30 ($e^{(-.036+.296)} = 1.297$) times greater than girls', the coefficient reaching the threshold of statistical significance ($p = .006$). The combination of national examinations and low between-teacher instructional variation is linked to a significant increase in gender equality in math scores and an increase in gender inequality, in favor of girls, in the odds of failure. National examinations are not linked to the most non-egalitarian outcome, the odds of excelling, but low between-teacher instructional variation is associated with a significant increase in gender equality in this aspect of achievement.

The coefficients of most measures of gender stratification do not reach statistical significance in the equations of the gender slope for all three outcomes, before and after inclusion of the educational standardization variables. The only exception is the coefficient of women's political empowerment in the equation of bottom 10%, which reaches the threshold of statistical significance: women's higher participation in politics is linked to an increase in girls' advantage in the odds of avoiding failure. This result differs from some previous results, which report that women's political activity moderates the gender gap also in math scores and in the upper parts of the distribution (Else-Quest et al.2010; Guiso et al., 2008; Penner, 2008; Riegle-Crumb, 2005). Regarding economic participation, our results resemble Riegle-Crumb's but differ from those of Penner, who found that economic participation enlarged the gender gap, and from those of Guiso et al. and of Baker and Perkins Jones (1993), who report a moderating effect of women's participation in the labor market. As noted, this difference may result from differences in the samples. Penner concentrated on Western countries, whereas the countries in our sample represent a variety of cultures. Guiso et al. did not limit the analysis to Western countries, but their sample did not include Middle Eastern Muslim countries. As shown in Table 1, in some Muslim countries girls have higher scores than boys, although they score lower than boys in most Western countries. This implies that the association between women's political and economic

[16] This implies that in a country which is average regarding gender stratification there is no gender difference in scores. When this country is also average in between-teacher instructional variation and does not use national examination, boys exhibit some advantage over girls.

[17] In unreported analyses, we checked the statistical significance of the gender slope for each outcome in countries that use national examinations.

participation and the gender gap in mathematics is not universal. The link between these indicators of gender stratification and educational outcomes, which is expected in Western culture, does not necessarily hold for other cultures. Some researchers, perhaps paradoxically, ascribe this pattern to the centrality of domestic roles for women in developing countries. An example is the Middle East, where well educated women have a low connection to the labor market due to cultural preferences for domestic roles. These women view their education as functional for the domestic and not the public sphere (Read and Oselin, 2008). The relevance of math for a future occupational career—a major rationale of the hypothesized link between the gender gap in math achievement and gender stratification—does not seem to hold in this context. Ma (2008) suggests that the stereotype of mathematics as a "masculine" area, which prevails in developed countries, does not hold in the developing world, where traditional gender stereotypes typically emphasize social role rather than academic ability. Consequently, when girls have the opportunity to study math they do not consider it as a threat; sometimes they do very well, as in Jordan and Bahrain, and outperform boys. Lenzner (2006) supports this claim, showing that female enrollment in mathematics and computer sciences in higher education is particularly high in Muslim countries. This line of reasoning is supported by Charles and Bradley (2009), who argued, as noted before, that in developed societies, where gender is central in establishing identity, and mathematics is viewed as a "masculine" domain, girls may be reluctant to take math and may avoid related programs.

A different angle is offered by Fryer and Levitt (2010), who propose that the relatively high math achievements of girls in Muslim countries may be a result of the prevalence of single-gender education. Although the effects of this practice for boys and girls are controversial (Wiseman, 2008), researchers agree that it increases gender-atypical patterns of course-taking (Lenzner, 2006; Sullivan et al., 2010). This suggests that single-sex schooling may enhance girls' exposure to mathematics, hence reduce the gender gap in mathematics achievement. In Bahrain and Jordan (and also in Iran) the rate of schools that have a gender-segregation is about 90% or more (Wiseman, 2008), providing some support for this line of explanation. Fryer and Levitt suggest, however, that further research is needed to substantiate this proposition.

Another possible explanation for girls' relatively high achievements in developing countries is selection: female education may be less developed in these countries and girls in the education system are a selective group. An (unreported) analysis with percent female examinees in each country showed no effect whatsoever of this variable on the gender slope in all three outcomes.[18] Similar results were obtained by Fryer and Levitt.

The difference between our results and some previous results may also stem from differences in respondents' ages. We analyze eighth graders, whereas Penner (who used TIMSS, 1995) analyzed students in the last year of secondary school. Gender stratification may be more significant for older students (Lippman, 2002). Still, Guiso et al. (who used PISA, 2003) analyzed 15-year-olds, a group close in age to our sample. Riegle-Crumb, who, like us, did not find an effect of women's economic activity but did find a moderating effect of women's political activity, analyzed students from seventh and eighth grades. Development had no effect on the slope in any of the three equations. This result is similar to that of Riegel-Crumb, of Fryer and Levitt, and of Guiso and colleagues.[19]

Although it is not our main topic, we briefly refer to the coefficients of the variables in the equation of the intercept. The intraclass correlation coefficient shows that between-country variance in mean math score is 41% of the total variance. Students in economically developed countries and in countries with higher level of women's economic participation win higher math scores. Higher rates of women's participation in politics and high fertility rates are linked to lower scores.[20] The educational structure variables have no effect on the mean level of achievement.

The intraclass correlation coefficients for top and bottom 10% are very low (.008 and .003, respectively). This is not surprising. These variables were a priori defined as the odds of being at the top or bottom of the distribution in each country, and the minor between-country differences stem only from rounding and have no substantial meaning. In hierarchical analysis, variables included in the equation of a slope should also be included in the equation of the intercept (Raudenbush and Bryk, 2002). We therefore include country's characteristics in the intercept's equations for these two outcomes, but we see no point in presenting or discussing the equations of these intercepts.

6. Conclusions

Comparative research shows promise in the search for the causes of the gender gap in mathematics achievement. Between-country comparisons allow analysis of how it is related to structures, policies, values and norms. Whereas previous

[18] Due to the small number of units we decided not to include percent female in the final models. However, percent female had no effect on the slope even when it was the only country variable in the analysis.

[19] In order to asses whether our major results were affected by the choice of countries we conducted parallel analyses on a sub-sample of 18 developed countries, all of them belong to the OECD. Hierarchical analysis on 18 units requires much caution in interpreting the results, and we only refer to the general pattern. The current results had the same general pattern as the reported ones regarding the educational standardization variables: between-teacher instructional variation was linked to increasing inequality, whereas national examinations were linked to its moderation. It is interesting to note that country's non-educational characteristics, particularly development, had now significant coefficients, with the latter linked to an increase of gender inequality for the three outcomes. This suggests that the effect of educational standardization on the gender gap in math may be less country-dependent than that of economic development and gender stratification.

[20] The negative association between women's political empowerment and mean math score, which may look counterintuitive, stems from the fact that some countries with a relatively high average score (Japan, Korea, Malaysia) have low rates of participation of women in politics, while countries with particularly low scores (South Africa, Philippines) have a relatively high level of participation of women in politics. Obviously the link is statistical, not causal.

research studied gender stratification and economic development as country characteristics that may explain gender inequality, we concentrated on structural features of national educational systems and studied level of standardization as a correlate of between-country differences in the gender gap in mathematics.

Based on Trends of International Mathematical and Science Study (TIMSS) 2003, we compared the association between two variables that represent level of standardization, the use of national examinations and between-teacher instructional variation in the time devoted to major math topics, and the gender gap in math achievement in 32 countries. Math achievement was defined according to scores, odds of excelling, and odds of failing.

Our main results revealed a substantial association of the standardization measures, which represent characteristics of national educational systems, with the gender gap. Higher uniformity in time devoted to various math topics is linked to reduced gender gap in all three educational outcomes: scores, the odds of excelling and the odds of failing. The use of national examinations is linked to a reduced gender gap in math scores, and in the odds of failing. So countries that govern the curriculum through national examinations and those that expose students to similar math knowledge are characterized by lower levels of gender inequality. The moderation of gender inequality is not accompanied by a lowering of standards. The standardization measures are not associated with a country's average achievement.

These results also suggest that level of standardization is more central than women's economic and political activity, the primacy of family in their lives, and the country's human and economic development in the shaping of gender inequality in math. The educational reality that students experience in their schools therefore seems more significant than signs and expressions of gender stratification prevalent in the society and probably also in school.

Coupled with previous research on the negative link between standardization and socioeconomic inequality in achievement, our results support claims by proponents of national examinations regarding their positive impact on equality in achievement. The clear standards, the transparency of educational structures and the universal rules in systems that use national examinations seem to be similarly related to two major sources of educational inequality: socioeconomic background and gender. So perhaps paradoxically, educational standardization, which may be viewed as a conservative practice that limits the freedom of students to choose courses that match their abilities and inclinations, seems to be linked to more egalitarian educational outcomes.

Although it is not our major research question, it is worthwhile to point that our results support the claim of recent research that average achievement may be a less appropriate measure of gender inequality than top and bottom percentiles (Penner, 2008; Xie and Shauman, 2003). Boys have higher odds of excelling in most of the 32 countries, whereas they have a modest advantage in scores in 12 countries only. Girls' odds of failing are significantly higher than boys' only in six countries. The gender gap in favor of boys is thus most prominent in the upper part of the achievement distribution, and does not exist in the lowest part. In other words, boys reach the highest math achievements more than girls do, but the odds of getting the lowest achievements are quite similar for the two genders.

We wish to offer some directions for future comparative research on gender inequality in mathematics. First and foremost, future research should refer more extensively to features of educational systems. With educational systems being a very common aspect of comparative research on the link between socioeconomic status and achievement, their marginality in research on gender inequality is surprising. Our results show that this line of research may open a new angle in the investigation of gender inequality in achievement.

Our study also shows the limitations of research on gender inequality concentrating on developed countries. As we have shown here, developing countries sometimes exhibit counterintuitive patterns of gender inequality (see also Fryer and Levitt, 2010; Grant and Behrman, 2010). The findings concerning Muslim countries such as Bahrain and Jordan suggest that the link between gender equality in a country and gender equality in math achievement does not necessarily follow our expectations. A quest for a better understanding of the causes and consequences of this phenomenon is in order.

Appendix A

Table A1. Variables and descriptive statistics.

Student variables Variable name	Description		
Quantitative variables		Mean	S.D.
Math scores	Internationally standardized with mean of 500 and standards deviation of 100	455.87	105.78
Dummy variables		Proportion of category 1	
Gender	1 – male; 0 – female	.48	
Top 10%	1 – in top 10% his/her country; 0 otherwise	.10	
Bottom 10%	1 – in bottom 10% in his/her country; 0 – otherwise.	.10	

Table A1. (*continued*)

Country variables Variable name	Description		
Quantitative Variables		Mean	S.D.
Human and economic development. Composed of:	Constructed by principal component analysis	.00	1.00
Urban population	Percent urban population in the country	70.12	15.23
Human Development Index (HDI)[a]	Combination of indicators of life expectancy, educational attainment (mean and expected years of schooling) and standard of living (GNI per capita)	.83	.11
Cronbach's alpha	.79		
Women's economic participation Composed of:	Constructed by principal component analysis	.00	1.00
Women's economic activity	Percent women aged 15 and older in the labor market	46.01	11.75
Women's participation in the labor market	Women's share in the labor force	39.57	8.75
Cronbach's alpha	.92		
Women's political empowerment Composed of:	Constructed by principal component analysis	.00	1.00
Women in parliament	% women in parliament	.19	1.11
Women in government	% women in government	.18	.12
Cronbach's alpha	.90		
Primacy of family Composed of:		.00	1.00
Total fertility rate	Number of children that would be born to each woman if she were to live to the end of her child-bearing years and bear children at each age in accordance with prevailing age-specific fertility rate	2.10	.83
Adolescent fertility rate	Number of births to women aged 15–19 per 1000 women aged 15–19	25.84	18.92
Cronbach's alpha	.73		
Between-teacher instructional variation. Composed of:	Constructed by principal component analysis	.00	1.00
Coefficients of variation of teachers' answers on time they devote to 5 math topics:	Categories of the original variables: 1 – never; 2 – some lessons; 3 – about half of the lessons; 4 –almost all lessons.		
Addition		.31	.05
Fractions		.27	.04
Problems		.31	.05
Data tables		.24	.06
Equations		.28	.04
Cronbach's alpha	.73		
Dummy variables		Proportion of category 1	
National examinations	1 – country uses national assessment to monitor implementation of national curriculum; 0 – otherwise	.53	

[a] Calculated by Human Development Report Office (HDRO) (http://hdr.undp.org/en/media/HDR_2011_EN_Tables.pdf).

Table A2. Correlations among the variables.

Student variables	Top 10%	Bottom 10%	Male
Scores	.414*	−.398*	.016*
Top 10%		−.111*	.029*
Bottom 10%			.018*

* $P < .05$.

Table A2. (*continued*)

Country variables	Economic participation	Political empowerment	Primacy of family	National examinations	Between-teacher variation
Human and economic development	.124	.388*	−.583*	.152	−.427*
Women's economic participation		.536*	−.123	−.073	−.489*
Women's political empowerment			−.144	−.030	−.377*
Primacy of family				−.062	.309**
National examinations					−.168

* $p < .05$.
** $p < .10$.

References

Allmendinger, J., 1989. Educational system and labor market outcomes. European Sociological Review 5, 231–250.
Al Mahadin, S., 2004. Jordanian women in education: politics, pedagogy and gender discourses. Feminist Review 78, 22–37.
Al-Nhara, T., 1999. Determinants of grade 8 achievement in Jordan: a multilevel analysis approach. Educational Psychology 19 (1), 37–52.
Amerin, A.L., Berliner, D.C., 2002. The Impact of High-Stakes Tests on Student Academic Performance: An Analysis of NAEP Results in States With High-Stakes Tests and ACT, SAT, and AP Test Results in States With High School Graduation Exams. Arizona State University: Education Policy Studies Laboratory Education Policy Research Unit.
Ayalon, H., Gamoran, A., 2000. Stratification in academic secondary programs and educational inequality in Israel and the United States. Comparative Education Review 44, 54–80.
Baker, D.P., Perkins Jones, D., 1993. Creating gender equality: cross-national gender gratification and mathematical performance. Sociology of Education 66, 91–103.
Carnoy, M., Loeb, S., 2002. Does External Accountability Affect Student Outcomes? A Cross-State Analysis. Educational Evaluation and Policy Analysis 24, 305–331.
Charles, M., Bradley, K., 2009. Indulging our gendered selves? Sex segregation by field of study in 44 countries. American Journal of Sociology 114, 924–976.
Davenport Jr., E.C., Davison, M.L., Kuang, H., Ding, S., Kim, S., Kwak, N.K., 1998. High school mathematics coursetaking by gender and ethnicity. American Education Research Journal 35, 497–514.
Diamond, J.B., 2007. Where the rubber meets the road: rethinking the connection between high-stakes testing policy and classroom instruction. Sociology of Education 80, 285–313.
Else-Quest, N.M., Hyde, J.S., Linn, M.C., 2010. Cross-national patterns of gender differences in mathematics: a meta-analysis. Psychological Bulletin 136, 102–127.
Fryer Jr., R.G., Levitt, S.D., 2010. An empirical analysis of the gender gap in mathematics. American Economics Journal: Applied Economics 2, 210–240.
Fuchs, T., Woessmann, L., 2007. What accounts for international differences in student performance? A reexamination using PISA data. Empirical Economic 32, 433–464.
Gamoran, A., Hannigan, E.C., 2000. Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. Educational Evaluation and Policy Analysis 22, 241–253.
Garden, R.A., 1989. Students' achievement: population B. In: Robitaille, D.F., Garden, R.A. (Eds.), The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics. Pergamon Press, Oxford, pp. 126–152.
Grant, M.J., Behrman, R.J., 2010. Gender gaps in educational attainment in less developed countries. Population and Development Review 36 (1), 71–89.
Guiso, L.F., Monte, P., Sapienza, P., Zingales, L., 2008. Culture, gender, and math. Science 320, 1164–1165.
Hyde, J.S., Mertz, J.E., 2009. Gender, culture, and mathematics performance. PNAS 106, 8801–8807, http://www.google.com/books?hl=iw&lr= amp;id=n4fAYzBcYEwC&oi=fnd&pg=PA17&dq=gender+gap+mathematics+Arabic+Muslim+countries&ots=_7GBN_DE-4&sig=rYB1KkMpaRGcZgB2Hj7ZgujHk1g#v=onepage&q&f=false.
Lenzner, A., 2006. Women in Mathematics: A Cross-Cultural Comparison. Waxmann Verlag GmbH, Munster.
Lippman, L., 2002. Cross-National variation in educational preparation for adulthood: from early adolescence to young adulthood. Annals of the American Academy of Political and Social Science 580, 70–102.
Loeb, S., Strunk, K., 2007. Accountability and local control: response to incentives with and without authority over resource generation and allocation. Educational Finance and Policy 2, 10–39.
Ma, X. 2008. Gender differences in mathematics achievement: Evidence from latest regional and international student assessments. In: Paper presented at the 11th International Congress on Mathematical Education. Monterrey, Mexico.
Mandel, H., Semyonov, M., 2005. Family policies, wage structures, and gender gaps: sources of earning inequality in 20 countries. American Sociological Review, 949–967.
McFarland, D.A., Rodan, S., 2009. Organization by design: supply- and demand-side models of mathematical course taking. Sociology of Education 82, 315–343.
Mehran, G., 2003. The paradox of tradition and modernity in female education in the Islamic republic of Iran. Comparative Education Review 47, 269–286.
Montt, G., 2011. Cross-national differences in educational achievement inequality. Sociology of Education 84, 49–68.
Muller, C., Schiller, K.S., 2000. Leveling the playing field? Students' educational attainment and state's performance testing. Sociology of Education 73, 196–218.
Park, H., 2008. The varied educational effects of parent-child communication: a comparative study of fourteen countries. Comparative Education Review 52, 219–243.

Penner, A., 2008. Gender differences in extreme mathematical achievement: an international perspective of biological and social forces. American Journal of Sociology 114 (Suppl.), S138–S170.
Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H., Rasbash, J., 1998. Weighting for unequal selection probabilities in multilevel models. JRSS Series B 60, 123–140.
Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical Linear Models, second edition. Sage Publications, Thousand Oaks.
Read, J.G., Oselin, S., 2008. Gender and the education-employment paradox in ethnic and religious contexts: the case of Arab Americans. American Sociological Review 73 (2), 296–313.
Resh, N., Benavot, A., 2009. Educational governance, school autonomy and curriculum implementation: diversity and uniformity in knowledge offerings to Israeli students. Journal of Curriculum Studies 41, 67–92.
Riegle-Crumb, C., 2005. The cross-national context of the gender gap in math and science. In: Hodges, L.V., Schneider, B. (Eds.), The Social Organization of Schooling. Russell Sage Foundation, New York, pp. 227–243.
Riegle-Crumb, C., 2006. The path through math: course sequences and academic performance at the intersection of race–ethnicity and gender. American Journal of Education 113, 101–122.
Stevenson, D.L., Baker, D.P., 1991. State control of the curriculum and classroom instruction. Sociology of Education 60, 104–122.
Sullivan, A., Joshi, H., Leonard, D., 2010. Single-sex schooling and academic attainment at school and through the lifecourse. American Educational Research Journal 47, 6–36.
TIMSS. <http://nces.ed.gov/timss/results03.asp> (accessed 15.04.07).
Tsui, M., 2007. Gender and mathematics achievement in China and the United States. Gender Issues 24, 1–11.
UNPD. <http://hdr.undp.org/statistics> (accessed 15.04.07).
Van de Werfhorst, H.G., Mijs, J.B., 2010. Achievement inequality and the institutional structure of educational systems: a comparative perspective. Annual Review of Sociology 36, 407–428.
Wiseman, A.W., 2008. A culture of (in)equality? A cross-national study of gender parity and gender segregation in national school systems. Research in Comparative and International Education 3, 179–201.
Watanabe, M., 2008. Tracking in the era of high-stakes state accountability reform: case studies of classroom instruction in North Carolina. Teachers College Record 110, 489–534.
Xie, Y., Shauman, K.A., 2003. Women in Science: Career Processes and Outcomes. Harvard University Press, Cambridge, MA.