# Preferences, Access, and the STEM Gender Gap in Centralized High School Assignment

Diana Ngo and Andrew Dustan

## Online Appendix

## A. Additional context

### A.1 Returns to STEM in Mexico

To estimate the labor market returns to STEM occupations, we use data from the third quarter of the 2010 and 2012 Encuesta Nacional de Ocupación y Empleo, ENOE, conducted by the Instituto Nacional de Estadística y Geografía, INEGI (INEGI 2012b). The ENOE is a nationally representative survey with information on employment, occupation, monthly income, and hours worked. In this same quarter, an additional module (Encuesta Nacional de Inserción Laboral de los Egresados de la Educación Media Superior, ENILEMS (INEGI 2012a)), collected information on recent high school graduates and their transition to higher education and/or the workforce, including their high school concentration, college major, and occupation, when applicable. Note, the ENILEMS survey respondents can report simultaneously enrolling in college and working.[1]

We use the same Brookings classifications described in the main text to classify occupations as STEM or non-STEM. Specifically, INEGI provides a crosswalk between the Mexican occupation codes and the U.S. Bureau of Labor Statistics Standard Occupation Classification (SOC) codes (INEGI 2012c), with each Mexican occupation matching one or more SOC codes. We compute the average STEM classification of all matched occupations and categorize the Mexican occupation/education track as STEM if the average STEM classification is 0.5 or more. Each occupation/education track is double-coded, and discrepancies are reconciled by a third individual. The resulting dataset is included in Ngo and Dustan (2023).

Using these data, we estimate the STEM wage premia for females as well as the transition probabilities between various levels of STEM education and STEM work. We restrict all analyses to the sample of respondents 40 years old or younger, who are more likely to represent the high

[1] Since the larger ENOE module only asks about terminal degrees, we are unable to fully trace individuals' pathways through high school tracks to college majors to occupations. The ENILEMS module allows us to see both high school tracks and college majors for the small subsample; however, since these are recent high school graduates, we do not observe their post-college careers.

school students in our context.

To determine whether or not the STEM wage premium remains after controlling for selection into STEM, we also include a set of basic demographics available in the data: urban, age, household headship, household size, and marital status. We also include parental education in a separate analysis. ENOE does not obtain information on parental education from all respondents, so we use education levels from the household roster and run the analysis for the subsample of respondents who are children of the household head. Since this is a selected sample, we prefer the full results.

## A.2 Commute Survey and Travel Time

To understand the context of commuting in Mexico City and to estimate travel times, we use the 2017 Encuesta Origen Destino en Hogares de Zona Metropolitana del Valle de Mexico, EOD (INEGI 2017). The EOD is a survey on travel in and around Mexico City, with detailed information on the purpose, travel modes, costs, and timing of trips. Although one can obtain travel times directly using OSRM, we use the EOD to more accurately account for traffic, transit delays, and transfers. We calculate distances for EOD trips using the OSRM distances between origin and destination district centroids. We then run cubic regression models of reported travel time on trip distance, including weekday trips made by high school age respondents (ages 15 to 18) for school (to school, from school, or for study purposes). We run separate models within each region (Mexico City, State of Mexico East, and State of Mexico West) and between region pairs. We do not have detailed distance information on trips within districts and exclude these. Finally, we use these results to calculate travel times in the COMIPEMS data.

We also use the EOD to summarize travel costs. The survey asks respondents to report costs for private and privately operated transit modes (e.g., micro, private buses). For driving trips, this includes the daily cost of parking but does not include the cost of gasoline. For public transit modes, we use the 2017 fares:

- Metro: 5 pesos
- Metrobus: 6 pesos (https://mexicocity.cdmx.gob.mx/e/getting-around/using-the-metro/)
- Tren ligero: 3 pesos (https://www.ste.cdmx.gob.mx/tren-ligero)
- Tren suburbano: 7 pesos for trips 0-12.8km, 16 pesos for trips 12.9-15.6km (https://www.elfinanciero.com.mx/nacional/tarifas-del-suburbano-aumentaran-a-partir-de-este-martes/)
- Trolebus: Between 2 and 5 pesos—using 2.5 pesos
- RTP/M1 buses: 2 pesos, though can be slightly more for express or eco buses (https://theculturetrip.com/north-america/mexico/articles/a-users-guide-to-the-mexico-c

ity-public-transport-system/, https://en.wikivoyage.org/wiki/Mexico_City, https://en.wikipedia.org/wiki/Red_de_Transporte_de_Pasajeros)

- Mexicable: 6 pesos (https://www.jornada.com.mx/2017/09/10/estados/028n2est)

Table A1: STEM wage differentials

Panel A. Full sample

| | Low education: High school or less | | | High education: College or more | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | ln(Wage) | ln(Wage) | ln(Wage) | ln(Wage) | ln(Wage) | ln(Wage) |
| STEM occupation | 0.09 | 0.07 | 0.06 | 0.18 | 0.15 | 0.14 |
| | (0.009) | (0.009) | (0.009) | (0.021) | (0.021) | (0.021) |
| Female X STEM occupation | 0.10 | 0.11 | 0.11 | -0.00 | -0.01 | 0.00 |
| | (0.022) | (0.022) | (0.022) | (0.031) | (0.030) | (0.030) |
| Urban | | 0.15 | 0.15 | | 0.07 | 0.06 |
| | | (0.005) | (0.005) | | (0.013) | (0.013) |
| Age | | 0.01 | 0.01 | | 0.04 | 0.03 |
| | | (0.000) | (0.000) | | (0.001) | (0.001) |
| Household head | | | 0.05 | | | 0.06 |
| | | | (0.007) | | | (0.018) |
| Household size | | | -0.01 | | | -0.04 |
| | | | (0.001) | | | (0.004) |
| Married | | | 0.07 | | | 0.11 |
| | | | (0.006) | | | (0.016) |
| Observations | 114899 | 114899 | 114899 | 28584 | 28584 | 28584 |
| Adjusted $R^2$ | 0.088 | 0.124 | 0.130 | 0.052 | 0.139 | 0.158 |
| Mean wage, non-STEM occupation, male | 23.00 | 23.00 | 23.00 | 43.30 | 43.30 | 43.30 |
| Mean wage, non-STEM occupation, female | 21.82 | 21.82 | 21.82 | 43.23 | 43.23 | 43.23 |
| Total effect, female STEM occupation | 0.18 | 0.17 | 0.18 | 0.18 | 0.14 | 0.14 |
| | (0.020) | (0.020) | (0.020) | (0.023) | (0.022) | (0.022) |

Panel B. Children of household head

| | Low education: High school or less | | | High education: College or more | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | ln(Wage) | ln(Wage) | ln(Wage) | ln(Wage) | ln(Wage) | ln(Wage) |
| STEM occupation | 0.06 | 0.04 | 0.04 | 0.17 | 0.13 | 0.12 |
| | (0.016) | (0.015) | (0.015) | (0.030) | (0.028) | (0.028) |
| Female X STEM occupation | 0.07 | 0.07 | 0.07 | 0.02 | 0.01 | 0.01 |
| | (0.033) | (0.032) | (0.032) | (0.045) | (0.042) | (0.041) |
| Urban | | 0.12 | 0.12 | | 0.08 | 0.07 |
| | | (0.008) | (0.008) | | (0.019) | (0.019) |
| Age | | 0.01 | 0.01 | | 0.04 | 0.04 |
| | | (0.001) | (0.001) | | (0.002) | (0.002) |
| High parental education (high school or more) | | 0.14 | 0.14 | | 0.23 | 0.22 |
| | | (0.013) | (0.013) | | (0.019) | (0.019) |
| Household size | | | -0.00 | | | -0.03 |
| | | | (0.002) | | | (0.005) |
| Married | | | 0.08 | | | 0.01 |
| | | | (0.014) | | | (0.038) |
| Observations | 44037 | 44037 | 44037 | 12034 | 12034 | 12034 |
| Adjusted $R^2$ | 0.087 | 0.124 | 0.125 | 0.056 | 0.149 | 0.155 |
| Mean wage, non-STEM occupation, male | 20.42 | 20.42 | 20.42 | 35.48 | 35.48 | 35.48 |
| Mean wage, non-STEM occupation, female | 19.23 | 19.23 | 19.23 | 36.31 | 36.31 | 36.31 |
| Total effect, female STEM occupation | 0.12 | 0.11 | 0.11 | 0.19 | 0.13 | 0.13 |
| | (0.029) | (0.028) | (0.028) | (0.033) | (0.031) | (0.031) |

Note: Sample is comprised of individuals aged 40 or younger from the 2010 and 2012 ENOE who report being paid on a regular schedule. Panel B includes the subset with a measure of parental education (i.e., children of the household head). All regressions include state by year by sex fixed effects. Huber-White robust standard errors are in parentheses.

Table A2: Transitions between STEM education and later STEM-related activities

| | ENILEMS module subsample | | | | Full ENOE sample |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | STEM college major | STEM job | STEM job | STEM job | STEM job |
| | | Unconditional | No college | In college | |
| STEM high school | 12.0 | 7.1 | 6.3 | 8.1 | |
| | (2.54) | (2.55) | (3.65) | (3.85) | |
| Female X STEM high school | 4.8 | -0.3 | 1.8 | -4.1 | |
| | (3.75) | (3.25) | (4.61) | (4.70) | |
| STEM college major | | | | | 33.7 |
| | | | | | (0.73) |
| Female X STEM college major | | | | | 6.8 |
| | | | | | (1.13) |
| Observations | 10197 | 5837 | 3032 | 2794 | 65997 |
| Adjusted $R^2$ | 0.072 | 0.051 | 0.062 | 0.074 | 0.171 |
| Dep. var. mean, non-STEM high school, male | 59.4 | 12.3 | 12.7 | 12.0 | |
| Dep. var. mean, non-STEM high school, female | 40.3 | 4.8 | 4.4 | 5.2 | |
| Dep. var. mean, non-STEM college, male | | | | | 11.8 |
| Dep. var. mean, non-STEM college, female | | | | | 7.5 |
| Total effect, female STEM high school | 16.8 | 6.7 | 8.2 | 4.0 | |
| | (2.76) | (2.02) | (2.81) | (2.70) | |
| Total effect, female STEM college | | | | | 40.5 |
| | | | | | (0.86) |

Note: Data are from 2010 and 2012. The samples for columns (1) through (4) are from the ENILEMS labor force insertion module collected from recent high school graduates. Column (1) includes individuals who transitioned from high school into higher education. Column (2) is the sample who transitioned from high school into working (including concurrent higher education). Of these, column (3) is the subset who do not report any college while column (4) is the subset who are working while in college. Column (5) data are from the full ENOE sample. All regressions include state by year by sex fixed effects. Huber-White robust standard errors are in parentheses.

Table A3: STEM school assignment summary statistics for 2005, before and after post-computerized assignment phase

| | (1)<br>Male | (2)<br>Female | (3)<br>Difference |
|---|---|---|---|
| **Panel A. Computerized assignment, conditional on assignment** | | | |
| STEM assigned program (elite or non-elite) | 34.9 | 23.2 | 11.7 |
| | (47.7) | (42.2) | (0.2) |
| Elite STEM assigned program | 11.0 | 4.4 | 6.6 |
| | (31.3) | (20.4) | (0.1) |
| Non-elite STEM assigned program | 23.9 | 18.9 | 5.0 |
| | (42.6) | (39.1) | (0.2) |
| Elite non-STEM assigned program | 19.7 | 21.2 | -1.4 |
| | (39.8) | (40.9) | (0.2) |
| Technical non-STEM assigned program | 12.3 | 15.6 | -3.2 |
| | (32.9) | (36.2) | (0.2) |
| Unassigned | 13.5 | 20.4 | -6.9 |
| | (34.2) | (40.3) | (0.2) |
| Observations | 118979 | 125260 | 244239 |
| **Panel B. Finalized assignment, conditional on assignment** | | | |
| STEM assigned program (elite or non-elite) | 35.6 | 24.3 | 11.2 |
| | (47.9) | (42.9) | (0.2) |
| Elite STEM assigned program | 10.0 | 3.8 | 6.2 |
| | (30.0) | (19.2) | (0.1) |
| Non-elite STEM assigned program | 25.5 | 20.5 | 5.0 |
| | (43.6) | (40.4) | (0.2) |
| Elite non-STEM assigned program | 18.0 | 18.6 | -0.6 |
| | (38.4) | (38.9) | (0.2) |
| Technical non-STEM assigned program | 12.7 | 16.7 | -4.1 |
| | (33.3) | (37.3) | (0.2) |
| Unassigned | 5.8 | 10.0 | -4.2 |
| | (23.4) | (30.0) | (0.1) |
| Observations | 118979 | 125260 | 244239 |

Note: Calculations in Panel A are for all students who were assigned to a program by the placement algorithm in the 2005 COMIPEMS cycle who resided within the COMIPEMS geographical boundary. Calculations in Panel B include students assigned either by the placement algorithm or during the post-assignment program selection process, in which unassigned students were able to choose a program that had not filled its quota (or remain unassigned). Indicator variables are percentages. Standard deviations are in parentheses in columns (1) and (2); standard errors are in parentheses in column (3). "Unassigned" is the proportion of the full sample that was unassigned to any program at the end of the either the computerized or finalized assignment, respectively. Observation counts correspond to the full sample, including students who are unassigned.

Table A4: Commute survey summary statistics

Panel A. All districts

| | (1)<br>Male | (2)<br>Female | (3)<br>Difference |
|---|---|---|---|
| Distance (km by driving) | 8.3 | 7.9 | 0.4 |
| | (9.65) | (9.32) | (0.15) |
| Travel time (minutes) | 44.4 | 42.9 | 1.5 |
| | (31.53) | (30.34) | (0.49) |
| Cost (2017 MXN) | 7.8 | 8.0 | -0.2 |
| | (9.49) | (10.60) | (0.16) |
| Total transit modes (excluding walking) | 0.9 | 0.9 | 0.0 |
| | (0.67) | (0.65) | (0.01) |
| **Travel mode, percent of sample** | | | |
| Walking only | 24.0 | 23.3 | 0.7 |
| | (42.71) | (42.26) | (0.68) |
| Car | 8.6 | 9.8 | -1.2 |
| | (28.04) | (29.74) | (0.46) |
| Micro | 51.7 | 52.7 | -1.0 |
| | (49.97) | (49.93) | (0.79) |
| Metro | 13.5 | 10.9 | 2.6 |
| | (34.22) | (31.19) | (0.52) |
| Metrobus | 5.4 | 3.9 | 1.5 |
| | (22.64) | (19.34) | (0.33) |
| Autobus | 3.2 | 3.0 | 0.2 |
| | (17.59) | (17.08) | (0.28) |
| Observations | 7911 | 7910 | 15821 |

Panel B. Interdistrict trips only

| | (1)<br>Male | (2)<br>Female | (3)<br>Difference |
|---|---|---|---|
| Distance (km by driving) | 12.5 | 12.1 | 0.5 |
| | (9.37) | (9.08) | (0.18) |
| Travel time (minutes) | 55.5 | 53.0 | 2.6 |
| | (32.19) | (31.14) | (0.62) |
| Cost (2017 MXN) | 10.0 | 10.0 | -0.1 |
| | (9.75) | (11.48) | (0.21) |
| Total transit modes (excluding walking) | 1.2 | 1.2 | 0.0 |
| | (0.60) | (0.59) | (0.01) |
| **Travel modes, percent of sample** | | | |
| Walking only | 7.2 | 7.9 | -0.7 |
| | (25.85) | (26.91) | (0.52) |
| Car | 9.9 | 11.6 | -1.8 |
| | (29.82) | (32.08) | (0.61) |
| Micro | 65.4 | 64.9 | 0.5 |
| | (47.56) | (47.73) | (0.94) |
| Metro | 20.2 | 16.3 | 4.0 |
| | (40.19) | (36.92) | (0.76) |
| Metrobus | 7.6 | 5.5 | 2.1 |
| | (26.47) | (22.77) | (0.48) |
| Autobus | 4.2 | 4.3 | -0.1 |
| | (20.02) | (20.28) | (0.40) |
| Observations | 5199 | 5170 | 10369 |

Note: Data are from the 2017 EOD for weekday trips made by respondents ages 15 to 18 for school (to school, from school, or for study purposes). Trip distances are calculated using the OSRM driving distance between the origin and destination district centroids. Trips with invalid origin or destination districts are excluded. Transit modes are not mutually exclusive. Other less commonly used transit modes (e.g., light rail, suburban trains, mototaxis) are excluded.

Table A5: Selection into COMIPEMS sample

| | (1)<br>Appears in<br>COMIPEMS | (2)<br>ENLACE 9 math<br>score, normalized | (3)<br>ENLACE 9 Spanish<br>score, normalized |
|---|---|---|---|
| Female | 0.01 | -0.06 | 0.25 |
| | (0.001) | (0.006) | (0.006) |
| Appears in COMIPEMS sample | | 0.23 | 0.21 |
| | | (0.005) | (0.005) |
| Female X appears in COMIPEMS | | -0.04 | -0.01 |
| | | (0.007) | (0.006) |
| Constant | 0.80 | 0.20 | 0.13 |
| | (0.001) | (0.004) | (0.004) |
| Observations | 562593 | 562593 | 562593 |

Note: Data are from 2007 and 2008 ENLACE 9 test takers who resided within the COMIPEMS geographical boundary. Huber-White robust standard errors are in parentheses.

## Table A6: Education track STEM mapping

| Program code | STEM classification | Program code | STEM classification |
|---|---|---|---|
| 001 Administración | 0 | 072 Mantenimiento automotriz | 0 |
| 002 Refrigeración y aire acondicionado | 1 | 073 Producción industrial | 0 |
| 003 Análisis y tecnología de alimentos | 1 | 074 Sistemas de impresión offset y serigrafía | 0 |
| 006 Computación | 1 | 075 Telecomunicaciones | 1 |
| 007 Computación fiscal contable | 1 | 076 Técnico en mecatrónica | 1 |
| 008 Comunicación | 0 | 077 Técnico en manufactura asistida por computadora | 1 |
| 009 Construcción | 0 | 078 Técnico en alimentos instituciones educativas | 0 |
| 010 Contabilidad | 1 | 203 Agencia de viajes | 0 |
| 011 Dietética | 0 | 208 Artes gráficas | 0 |
| 012 Arquitectura | 1 | 214 Contabilidad | 0 |
| 013 Diseño gráfico | 0 | 218 Cosmetología esteticista | 0 |
| 014 Diseño de modas | 0 | 220 Dibujo publicitario | 0 |
| 015 Electricidad | 1 | 222 Diseño arquitectónico | 1 |
| 016 Electrónica | 1 | 223 Diseño decorativo | 0 |
| 017 Enfermería general | 1 | 224 Diseño gráfico | 0 |
| 018 Gericultura | 0 | 225 Diseño industrial | 1 |
| 019 Informática administrativa | 0 | 226 Diseño industrial de patrones | 1 |
| 020 Laboratorista clínico | 1 | 227 Ediciones | 0 |
| 021 Laboratorista químico | 1 | 229 Electricidad industrial | 1 |
| 022 Mantenimiento | 0 | 237 Fotomecánica | 0 |
| 023 Mantenimiento de equipo de computo | 1 | 238 Gerencia y supervisión en la industria del vestido | 0 |
| 024 Máquinas de combustión interna | 0 | 246 Mecánica automotriz | 0 |
| 025 Máquinas-herramienta | 1 | 247 Mecánica industrial | 1 |
| 026 Mecánica industrial | 1 | 250 Modelismo y fundición | 0 |
| 027 Producción | 0 | 252 Paquetes de cómputo | 1 |
| 028 Programador | 1 | 260 Radiología e imagen | 1 |
| 029 Prótesis dental | 1 | 264 Sastrería industrial | 0 |
| 030 Puericultura | 0 | 265 Secretario bilingue | 0 |
| 031 Secretario ejecutivo | 0 | 266 Secretario ejecutivo | 0 |
| 032 Supervisor en la industria del vestido | 0 | 267 Servicio a equipo de cómputo | 1 |
| 033 Técnico en agroindustrias | 1 | 275 Telecomunicaciones | 1 |
| 034 Técnico agropecuario | 1 | 277 Trabajo social | 0 |
| 035 Técnico en instrumentación dental | 1 | 278 Secretario ejecutivo bilingue | 0 |
| 036 Técnico en administración | 0 | 301 Administración | 0 |
| 037 Técnico en computacion fiscal contable | 1 | 302 Alimentos y bebidas | 0 |
| 038 Técnico en edificación | 1 | 303 Asistente directivo | 0 |
| 039 Técnico en contabilidad | 1 | 304 Automotriz | 0 |
| 040 Técnico en diseño industrial | 1 | 305 Construcción | 0 |
| 041 Técnico en diseño gráfico | 0 | 306 Contaduría | 1 |
| 042 Técnico en electricidad | 1 | 307 Control de calidad | 1 |
| 043 Técnico en electronica | 1 | 308 Conservación del medio ambiente | 1 |
| 044 Técnico en enfermería general | 1 | 309 Dental | 1 |
| 045 Técnico en industrializacion de lacteos | 1 | 310 Electricidad industrial | 1 |
| 046 Técnico en informática | 1 | 311 Electromecánica | 1 |
| 047 Técnico en informática agropecuaria | 1 | 312 Electrónica industrial | 1 |
| 048 Técnico en mantenimiento en equipo de computo | 1 | 313 Enfermería general | 1 |
| 049 Técnico en mantenimiento industrial | 1 | 314 Hospitalidad turística | 0 |
| 050 Técnico en maquinas-herramienta | 1 | 315 Industria del vestido | 0 |
| 052 Técnico laboratorista clinico | 1 | 316 Informática | 1 |
| 053 Técnico laboratorista químico-clínico | 1 | 317 Mantenimiento de equipo de cómputo y control digital | 1 |
| 054 Técnico en manufactura en la industria del vestido | 0 | 318 Mantenimiento de motores y planeadores | 1 |
| 055 Trabajo social | 0 | 319 Mantenimiento de sistemas automáticos | 1 |
| 056 Turismo | 0 | 320 Máquinas herramienta | 1 |
| 057 Técnico programador | 1 | 321 Metalmecánica | 0 |
| 058 Diseño decorativo | 0 | 322 Optometría | 1 |
| 059 Diseño industrial | 1 | 323 Plásticos | 0 |
| 060 Mecatrónica | 1 | 324 Procesamiento industrial de alimentos | 0 |
| 061 Técnico en horticultura | 1 | 325 Producción y transformación de productos acuícolas | 0 |
| 062 Técnico en sistemas electricos de control y automatizacion | 1 | 326 Productividad industrial | 0 |
| 063 Técnico asistente ejecutivo | 0 | 327 Química industrial | 1 |
| 064 Diseño y proyecto gráfico | 0 | 328 Refrigeración y aire acondicionado | 1 |
| 065 Asistente ejecutivo bilingüe | 0 | 329 Sistemas electrónicos de aviación | 1 |
| 066 Técnico en diseño asistido por computadora | 1 | 330 Telecomunicaciones | 1 |
| 067 Mantenimiento de equipo y sistemas | 0 | 331 Terapia respiratoria | 1 |
| 068 Informática | 1 | 332 Laministería y recubrimiento de las aeronaves | 0 |
| 069 Técnico en turismo | 0 | 333 Seguridad e higiene y Protección civil | 1 |
| 070 Técnico en gastronomía | 0 | 334 Expresión gráfica digital | 0 |
| 071 Técnico en mercadotecnia | 0 | 335 Mecatrónica | 1 |
| | | 336 Autotrónica | 1 |

Note: The guidelines for STEM classification come from Rothwell (2013), which identifies U.S. STEM occupations based on level of STEM knowledge required.

Table A7: Relationship between placement test score and ENLACE 9 subscores

| | Raw placement test score | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| ENLACE 9 math subscore (normalized) | 9.63 | | 9.05 |
| | (0.026) | | (0.026) |
| ENLACE 9 Spanish subscore (normalized) | 8.55 | | 9.28 |
| | (0.030) | | (0.030) |
| Male | | 3.99 | 4.64 |
| | | (0.065) | (0.039) |
| 2008 cohort | 1.32 | 1.26 | 1.40 |
| | (0.038) | (0.064) | (0.038) |
| Constant | 56.60 | 63.96 | 54.29 |
| | (0.030) | (0.054) | (0.034) |
| Observations | 373850 | 373850 | 373850 |
| Adjusted $R^2$ | 0.647 | 0.011 | 0.660 |
| Mean placement test score | 66.44 | 66.44 | 66.44 |

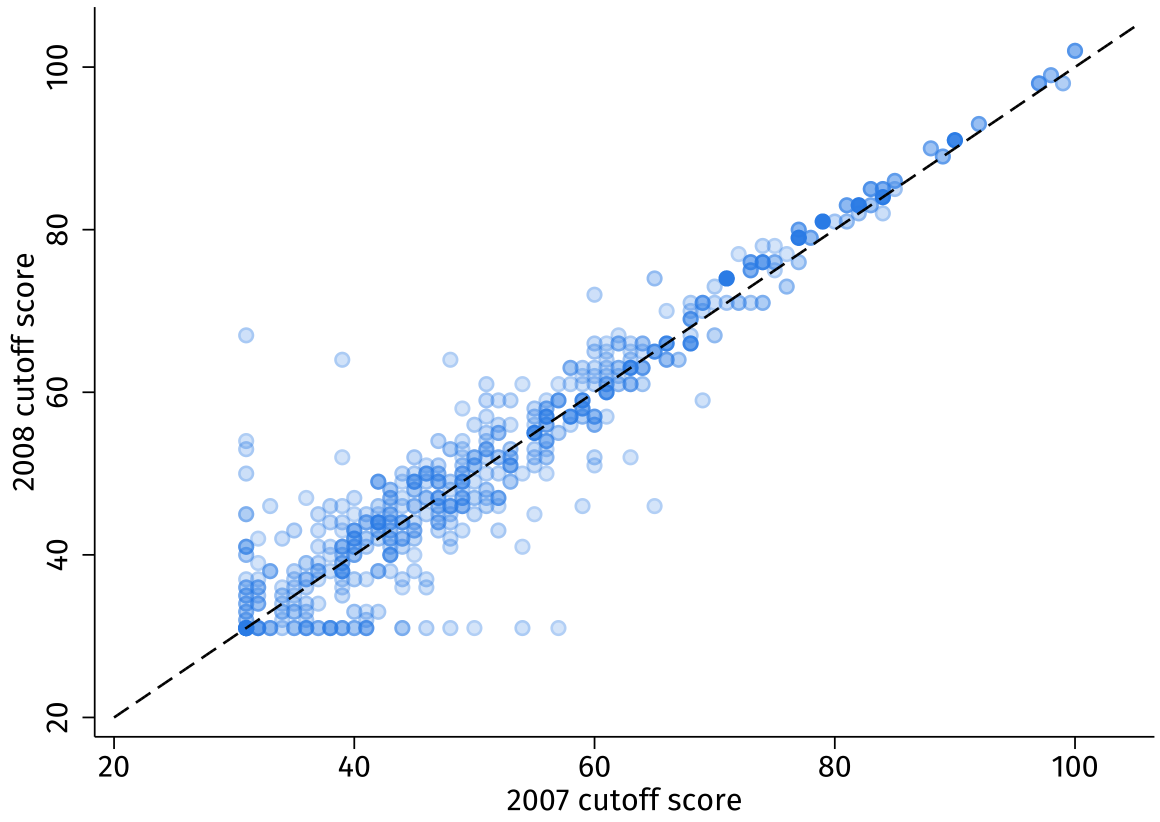
Note: Sample is the subset of the analysis sample that has ENLACE 9 scores available. Huber-White robust standard errors are in parentheses.

## Table A8: Payoff relevant “mistakes” and correlates

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Any mistake | Any mistake | STEM switching mistake | STEM switching mistake | Pro-STEM mistake | Pro-STEM mistake | Anti-STEM mistake | Anti-STEM mistake |
| Male | 1.3 | 0.7 | 1.2 | 0.5 | 1.1 | 0.5 | 0.1 | -0.0 |
| | (0.09) | (0.09) | (0.06) | (0.06) | (0.05) | (0.05) | (0.03) | (0.03) |
| High middle school GPA | -1.9 | -1.8 | -0.6 | -0.6 | -0.4 | -0.4 | -0.2 | -0.2 |
| | (0.10) | (0.10) | (0.07) | (0.06) | (0.06) | (0.06) | (0.03) | (0.03) |
| High ENLACE 9 math subscore | -0.5 | -0.5 | -0.0 | -0.1 | -0.0 | -0.1 | 0.0 | 0.0 |
| | (0.12) | (0.12) | (0.08) | (0.08) | (0.07) | (0.07) | (0.04) | (0.04) |
| High ENLACE 9 Spanish subscore | -0.6 | -0.5 | -0.5 | -0.4 | -0.4 | -0.3 | -0.1 | -0.1 |
| | (0.12) | (0.12) | (0.08) | (0.08) | (0.07) | (0.07) | (0.04) | (0.04) |
| Missing ENLACE 9 score | -0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.1 |
| | (0.25) | (0.25) | (0.17) | (0.16) | (0.15) | (0.15) | (0.08) | (0.08) |
| High parental education | -2.0 | -1.9 | -1.1 | -0.9 | -0.9 | -0.7 | -0.2 | -0.2 |
| | (0.10) | (0.10) | (0.06) | (0.06) | (0.06) | (0.06) | (0.03) | (0.03) |
| Missing parental education | -0.2 | -0.2 | -0.4 | -0.3 | -0.3 | -0.2 | -0.1 | -0.1 |
| | (0.14) | (0.14) | (0.09) | (0.09) | (0.08) | (0.08) | (0.05) | (0.05) |
| Middle school graduate | -0.1 | -0.1 | 0.0 | 0.1 | 0.1 | 0.2 | -0.1 | -0.1 |
| | (0.25) | (0.25) | (0.17) | (0.17) | (0.15) | (0.15) | (0.08) | (0.08) |
| Fraction of portfolio that is STEM | | 4.4 | | 12.5 | | 6.7 | | 5.7 |
| | | (1.00) | | (0.67) | | (0.59) | | (0.33) |
| Fraction of portfolio that is STEM$^2$ | | 21.0 | | 23.5 | | 27.7 | | -4.2 |
| | | (2.74) | | (1.82) | | (1.62) | | (0.90) |
| Fraction of portfolio that is STEM$^3$ | | -22.8 | | -35.0 | | -33.7 | | -1.3 |
| | | (1.87) | | (1.24) | | (1.10) | | (0.62) |
| Observations | 424688 | 424688 | 424688 | 424688 | 424688 | 424688 | 424688 | 424688 |
| Adjusted $R^2$ | 0.069 | 0.073 | 0.031 | 0.055 | 0.024 | 0.044 | 0.007 | 0.011 |
| Dep. var. mean | 9.1 | 9.1 | 3.7 | 3.7 | 2.9 | 2.9 | 0.8 | 0.8 |

Note: Sample is comprised of students from the analytical sample who were assigned. We identify each student’s “correct” assignment under no mistakes, i.e., the program they would have been assigned to if their portfolios were ordered by descending cutoffs. Mistakes are defined as having a “correct” assigned program cutoff that is one standard deviation (20 points) higher than the actual assigned program cutoff. STEM-switching mistakes are the subset that would have changed assignment from STEM to non-STEM or vice-versa. Pro-STEM mistakes are the subset where students were actually assigned to a STEM program when their “correct” assigned program was a non-STEM program, and anti-STEM mistakes are the subset where students were actually assigned to a non-STEM program when their “correct” assigned program was a STEM program. Huber-White robust standard errors are in parentheses.

Figure A1: Comparison of program-level cutoff scores in 2007 and 2008



Note: Markers correspond to program-level cutoff scores in 2007 and 2008. Cutoff scores are the lowest placement test score a student could obtain and be assigned, and are set to 31 (the minimum to be eligible for assignment) for programs that are not oversubscribed in that year. Opacity is determined by 2007 enrollment counts, such that darker points indicate higher enrollment. Dashed line is a 45-degree line. The raw correlation is 0.95 and enrollment-weighted correlation is 0.98.

# B. Decomposition and simulation details

## B.1 Decomposition

The decomposition exercise implements a version of Fairlie (2017) that decomposes STEM assignment probabilities conditional on any assignment (i.e. not remaining unassigned by the mechanism) and accounts for the covariate cell-based structure of the data.

Denote probability of assignment to program $j$, conditional on gender and all other observable characteristics, by $P\left(A = j|M, \tilde{X}\right)$, where $A = 0$ denotes that a student is unassigned by the mechanism. Assignment to a STEM program is denoted $S = 1$, i.e. $A \in \mathcal{S}$, where $\mathcal{S}$ is the set of STEM programs.

The gender gap in the the probability of STEM assignment conditional on any assignment can be decomposed as follows:

$$
\begin{aligned}
P\left(S = 1|A \neq 0; M = 1\right) - P\left(S = 1|A \neq 0; M = 0\right) &= \\
\frac{P\left(S = 1|M = 1\right)}{P\left(A \neq 0|M = 1\right)} - \frac{P\left(S = 1|M = 0\right)}{P\left(A \neq 0|M = 0\right)} &= \\
\frac{\int P(S = 1|M = 1, \tilde{X})dF(\tilde{X}|M = 1)}{\int P(A \neq 0|M = 1, \tilde{X})dF(\tilde{X}|M = 1)} - \frac{\int P(S = 1|M = 0, \tilde{X})dF(\tilde{X}|M = 0)}{\int P(A \neq 0|M = 0, \tilde{X})dF(\tilde{X}|M = 0)} &= \\
\underbrace{\frac{\int P(S = 1|M = 1, \tilde{X})dF(\tilde{X}|M = 1)}{\int P(A \neq 0|M = 1, \tilde{X})dF(\tilde{X}|M = 1)} - \frac{\int P(S = 1|M = 1, \tilde{X})dF(\tilde{X}|M = 0)}{\int P(A \neq 0|M = 1, \tilde{X})dF(\tilde{X}|M = 0)}}_{\text{Characteristic component}} &+ \\
\underbrace{\frac{\int P(S = 1|M = 1, \tilde{X})dF(\tilde{X}|M = 0)}{\int P(A \neq 0|M = 1, \tilde{X})dF(\tilde{X}|M = 0)} - \frac{\int P(S = 1|M = 0, \tilde{X})dF(\tilde{X}|M = 0)}{\int P(A \neq 0|M = 0, \tilde{X})dF(\tilde{X}|M = 0)}}_{\text{Preference component}} &.
\end{aligned}
$$

We note that $P\left(S = 1|M, \tilde{X}\right) = P\left(S = 1|A \neq 0; M, \tilde{X}\right) \cdot P\left(A \neq 0|M, \tilde{X}\right)$, so students contribute to the STEM gap both through their overall probability of assignment and their probability of STEM assignment conditional on any assignment.

We compute the sample analogues of the four terms of this decomposition in the following way. The first term can be computed from the relevant sample proportions or, equivalently, the mean predicted probabilities from the estimated conditional logit (suppressing region and year subscripts):

$$\frac{\int \hat{P}(S=1|M=1,\tilde{X})dF(\tilde{X}|M=1)}{\int \hat{P}(A\neq 0|M=1,\tilde{X})dF(\tilde{X}|M=1)} =$$

$$\frac{\bar{S}_{M=1}}{\overline{\mathbb{1}(A\neq 0)}_{M=1}} = \frac{\frac{1}{N_{M=1}}\sum_{i:M_i=1}(\sum_{j\in\mathcal{S}_i}\hat{V}_{ij}/\sum_{k\in\mathcal{J}_i}\hat{V}_{ik})}{\frac{1}{N_{M=1}}\sum_{i:M_i=1}(\sum_{\ell\in\mathcal{J}_i\setminus 0}\hat{V}_{i\ell}/\sum_{k\in\mathcal{J}_i}\hat{V}_{ik})},$$

where $\mathcal{S}_i$ is the feasible set of STEM programs for student $i$. The final term is analogous, but with $M = 0$ instead of $M = 1$.

The interior terms are the same as each other, and can be computed from the appropriate mean predicted probabilities, with $\hat{V}_{ij}(\hat{\beta}_{c(X_i,M=1)})$ indicating that the male covariate cell parameters (holding all other student covariates the same at $X_i$) are used in computing the observable utility component:

$$\frac{\int \hat{P}(S=1|M=1,\tilde{X})dF(\tilde{X}|M=0)}{\int \hat{P}(A\neq 0|M=1,\tilde{X})dF(\tilde{X}|M=0)} =$$

$$\frac{\frac{1}{N_{M=0}}\sum_{i:M_i=0}\left[\sum_{j\in\mathcal{S}_i}\hat{V}_{ij}(\hat{\beta}_{c(X_i,M=1)})/\sum_{k\in\mathcal{J}_i}\hat{V}_{ik}(\hat{\beta}_{c(X_i,M=1)})\right]}{\frac{1}{N_{M=0}}\sum_{i:M_i=0}\left[\sum_{\ell\in\mathcal{J}_i\setminus 0}\hat{V}_{i\ell}(\hat{\beta}_{c(X_i,M=1)})/\sum_{k\in\mathcal{J}_i}\hat{V}_{ik}(\hat{\beta}_{c(X_i,M=1)})\right]}.$$

These terms are sufficient for the aggregate decomposition into characteristic and preference components. The detailed decomposition is more complex. We follow the repeated simulation procedure in Fairlie (2017) to obtain the estimated contribution of each characteristic in a way that is invariant to the order in which the characteristics are "switched" from female to male. For each simulation repetition $r$, we:

1. Randomly determine the order in which female covariates will be switched to those of the matched males.

2. For each female in the sample, draw a male in the same region-year. Perform these draws with replacement.

3. In the random order determined in step 1, for each student characteristic indexed by $q$:

    - For each female, replace characteristic $q$ with the characteristic from her matched male. This is a cumulative replacement, so that any previously replaced characteristic is not reverted. Call the new vector of student characteristics $\tilde{X}_i(q)$. We switch the high parental education and missing parental education covariates at the same time.

    - Some $\tilde{X}_i(q)$ vectors correspond to an empty covariate cell $c(X_i(q), M = 1)$, such that assignment probabilities cannot be computed. This is the case when ENLACE 9 math score is set to missing while ENLACE 9 Spanish is set to non-missing, or vice versa. To prevent assignment to empty cells, we instead treat both ENLACE 9 scores as missing whenever one is missing. Similarly, middle school graduates always have missing ENLACE 9 scores, so we enforce missing ENLACE 9 as long as $\tilde{X}_i(q)$ has middle school graduate status.

- Compute the estimated probability of STEM assignment conditional on any assignment for females, with male preferences and the new student characteristics: $\frac{\int \hat{P}(S=1|M=1,\tilde{X})dF(\tilde{X}(q)|M=0)}{\int \hat{P}(A\neq 0|M=1,\tilde{X})dF(\tilde{X}(q)|M=0)}$.
- Compute the difference between the above estimated probability and the estimated probability obtained in the previous iteration. If this is the first iteration, compare to the estimated probability under the true covariate values $\tilde{X}$. This is the simulated contribution of characteristic $q$ to the STEM gender gap in repetition $r$, but it is dependent on the order in which the characteristics were replaced and the random draws from the male sample.

For each characteristic $q$, there are now $R = 100$ simulated contributions to the gap. We compute the mean of these contributions to obtain the contribution of $q$ while averaging over replacement orderings and male sample draws.

To account for uncertainty from the estimated preference parameters, we bootstrap standard errors for the aggregate and detailed decompositions. We perform $B = 50$ draws from the joint normal distribution of preference model parameters. In each simulation repetition $r$ described above, we compute the estimated preference and characteristic contributions for all $B$ parameter draws (in addition to the original estimated parameters). The standard deviation of these estimates over the parameter draws is the standard error of the contribution.

## B.2 Simulation

The simulation exercise proceeds as follows. We suppress region and year subscripts for clarity. Simulations are carried out separately by year, but pool all regions. We take the student populations as fixed (i.e. we do not resample students for the simulation). For each simulation repetition $r$, we:

1. Set program capacities. We set capacities (seat counts) to match those used in the respective year's actual assignment process. Capacities are unobserved for programs that did not fill up. We assume unlimited capacities for these programs, but results are similar if we fix capacities at the number of seats that were filled in that year.

2. Draw a single random tiebreaker $T_{ir}$ for each student. While in the true COMIPEMS assignment process, ties are resolved by school system representatives in real-time, who either

accept all tied applicants (exceeding capacity) or reject them all (leaving excess capacity), we keep the capacities fixed and use the tiebreaker.

3. Draw preference parameters $\{\boldsymbol{\delta}_r, \boldsymbol{\beta}_r\}$ and the scale parameter for $\eta$ from the joint normal distribution resulting from the maximum likelihood estimation of the conditional logit model.

4. Draw unobserved tastes $\widehat{\eta}_{ijr}$ from the appropriately scaled i.i.d. extreme value type I distribution.

5. For each counterfactual scenario $\ell$ (including the status quo):

    - Impose the counterfactual preferences (e.g. females have preferences of males with the same covariates), placement test scores (e.g. females have draws from the male conditional distribution), or priority structure (e.g. STEM programs add points to female placement test scores to determine priority rankings over students).

    - For each student, rank all programs by simulated utilities $\widehat{U}_{ijr\ell} = \widehat{V}_{ijr\ell} + \widehat{\eta}_{ijr}$. Call these ranked pseudoportfolios $R_{ir\ell}$. For this step, the "programs" that aggregate far-away alternatives are replaced with a randomly drawn feasible alternative among those that were aggregated.

    - For students whose preferences cannot be estimated—those outside the estimation sample because they are missing location information, are in the adult applicant category, or are outside the COMIPEMS zone—set $R_{ir\ell}$ equal to their actual, submitted portfolios.

    - Submit $\{\boldsymbol{R}_{r\ell}, \boldsymbol{T}_r\}$ to the deferred acceptance assignment mechanism and record assignments $A_{r\ell}$. Students who remain unassigned by the mechanism are denoted by $A_{ir\ell} = 0$.

    - For the estimation sample students only, compute proportions of assigned students who were assigned to each type of program. For example, to obtain the proportion of females assigned to STEM programs (collected in set $\mathcal{S}$), compute $\overline{\text{STEM}}_{r\ell}^{M=0} = \frac{1}{\sum_{i:M_i=0} \mathbb{1}(A_{ir\ell} \neq 0)} \sum_{\substack{i:M_i=0,\\ A_{ir\ell}\neq 0}} \mathbb{1}(A_{ir\ell} \in \mathcal{S})$. Store these proportions and their differences by gender.

6. Compute and store simulation repetition-specific differences in proportions of assigned students between counterfactual scenarios.

We perform 100 simulation repetitions and report the means and standard deviations of the relevant proportions. The standard deviations reflect uncertainty about preference parameters as well as simulation error introduced by the idiosyncratic preference and tiebreaker draws. This approach is similar to Pathak and Shi (2021), except that student characteristics are treated as fixed rather than resampling them in each simulation.

# C. Additional model and simulation results

Table C1: Model fit: simulated and actual gender gaps in choices and assignments

| | Simulated | | | Actual | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Male | Female | Difference | Male | Female | Difference |
| STEM | 38.1 | 27.1 | 11.1 | 38.7 | 27.6 | 11.1 |
| | (0.10) | (0.11) | (0.18) | | | |
| Elite STEM | 10.2 | 4.5 | 5.6 | 10.2 | 4.6 | 5.6 |
| | (0.04) | (0.04) | (0.09) | | | |
| Non-elite STEM | 28.0 | 22.5 | 5.5 | 28.4 | 23.0 | 5.5 |
| | (0.10) | (0.10) | (0.17) | | | |
| Elite non-STEM | 17.1 | 18.3 | -1.1 | 17.2 | 18.5 | -1.2 |
| | (0.06) | (0.06) | (0.11) | | | |
| Technical non-STEM | 12.7 | 15.5 | -2.8 | 12.8 | 15.7 | -2.9 |
| | (0.08) | (0.09) | (0.14) | | | |
| Traditional academic | 32.1 | 39.2 | -7.1 | 31.3 | 38.2 | -6.9 |
| | (0.09) | (0.09) | (0.17) | | | |

Note: Columns 1 through 3 report the simulated gender-specific proportions of assigned students who were assigned to the indicated program type, and their difference. Proportions are means over 100 independent simulations of the assignment process accounting for uncertainty in student preference parameters, idiosyncratic student preferences, and random tie-breakers in assignment. Standard deviations of the simulated proportions are in parentheses. Columns 4 through 6 show the actual proportions in the data. Proportions are reported in percentages. Simulations are as described in Section III.C, using estimated student preferences from the procedure described in Section III.A. Sample is 2007 and 2008 COMIPEMS cycles.

Table C2: Gender differences in preferences, by region

| | (1) | (2) | (3) |
|---|---|---|---|
| Elite STEM | 4.40 | 5.74 | 5.15 |
| | (0.197) | (0.319) | (0.464) |
| Non-elite STEM | 2.25 | 2.74 | 3.27 |
| | (0.079) | (0.077) | (0.125) |
| Elite non-STEM | -1.29 | -0.35 | -1.21 |
| | (0.231) | (0.488) | (0.404) |
| Technical non-STEM | -1.24 | 1.38 | 1.89 |
| | (0.091) | (0.095) | (0.142) |
| Unassigned | -0.92 | 0.25 | 0.08 |
| | (0.095) | (0.119) | (0.193) |
| Distance | -0.02 | 0.05 | 0.05 |
| | (0.004) | (0.004) | (0.005) |
| Region | Federal District | East SoM | West SoM |
| Proportion female (%) | 51.5 | 51.8 | 52.2 |

Note: Entries are estimated differences between male and female students in mean marginal utilities from the indicated program characteristics in the specified region of the COMIPEMS area, following equation 5 in Section IV.A. Driving distance is used in estimating the preference model. Standard errors are in parentheses.

Table C3: Program type preferences with respect to student characteristics

Panel A. Elite STEM programs

| | (1)<br>Male | (2)<br>Female | (3)<br>Difference |
|---|---|---|---|
| Parental education | 1.35 | 1.00 | 0.35 |
| | (0.179) | (0.309) | (0.357) |
| ENLACE 9 math subscore | 1.57 | 1.44 | 0.13 |
| | (0.211) | (0.359) | (0.416) |
| ENLACE 9 Spanish subscore | -0.81 | -0.84 | 0.03 |
| | (0.211) | (0.334) | (0.394) |
| Middle school GPA | 1.74 | 1.87 | -0.13 |
| | (0.187) | (0.312) | (0.364) |
| Middle school graduate | -5.02 | -1.02 | -4.00 |
| | (0.207) | (0.306) | (0.369) |

Panel B. Non-elite STEM programs

| | (1)<br>Male | (2)<br>Female | (3)<br>Difference |
|---|---|---|---|
| Parental education | -1.01 | -1.57 | 0.56 |
| | (0.079) | (0.073) | (0.107) |
| ENLACE 9 math subscore | 0.42 | 0.18 | 0.24 |
| | (0.106) | (0.092) | (0.140) |
| ENLACE 9 Spanish subscore | -0.10 | -0.50 | 0.41 |
| | (0.099) | (0.090) | (0.133) |
| Middle school GPA | -0.52 | -0.35 | -0.17 |
| | (0.080) | (0.080) | (0.113) |
| Middle school graduate | -1.06 | 0.44 | -1.51 |
| | (0.105) | (0.101) | (0.146) |

Panel C. Elite non-STEM programs

| | (1) | (2) | (3) |
|---|---|---|---|
| | Male | Female | Difference |
| Parental education | 2.10 | 2.10 | 0.00 |
| | (0.226) | (0.388) | (0.449) |
| ENLACE 9 math subscore | 0.87 | 0.68 | 0.19 |
| | (0.264) | (0.419) | (0.495) |
| ENLACE 9 Spanish subscore | 0.19 | -1.00 | 1.19 |
| | (0.263) | (0.404) | (0.482) |
| Middle school GPA | 1.84 | 2.71 | -0.88 |
| | (0.247) | (0.392) | (0.463) |
| Middle school graduate | -2.42 | -0.72 | -1.70 |
| | (0.229) | (0.354) | (0.421) |

Panel D. Technical non-STEM programs

| | (1) | (2) | (3) |
|---|---|---|---|
| | Male | Female | Difference |
| Parental education | -1.47 | -1.08 | -0.39 |
| | (0.104) | (0.080) | (0.131) |
| ENLACE 9 math subscore | 0.00 | 0.23 | -0.22 |
| | (0.134) | (0.102) | (0.168) |
| ENLACE 9 Spanish subscore | 0.27 | 0.11 | 0.17 |
| | (0.130) | (0.100) | (0.163) |
| Middle school GPA | -0.71 | -0.67 | -0.04 |
| | (0.103) | (0.087) | (0.135) |
| Middle school graduate | -1.49 | -0.19 | -1.30 |
| | (0.140) | (0.116) | (0.182) |

Panel E. Unassigned

| | (1)<br>Male | (2)<br>Female | (3)<br>Difference |
|---|---|---|---|
| Parental education | 3.54 | 3.36 | 0.17 |
| | (0.125) | (0.099) | (0.156) |
| ENLACE 9 math subscore | 0.19 | 0.46 | -0.27 |
| | (0.159) | (0.125) | (0.201) |
| ENLACE 9 Spanish subscore | 0.48 | 0.36 | 0.11 |
| | (0.154) | (0.119) | (0.194) |
| Middle school GPA | 1.56 | 0.60 | 0.96 |
| | (0.121) | (0.104) | (0.159) |
| Middle school graduate | 1.91 | 2.82 | -0.91 |
| | (0.147) | (0.127) | (0.193) |

Panel F. Distance

| | (1)<br>Male | (2)<br>Female | (3)<br>Difference |
|---|---|---|---|
| Parental education | 0.05 | 0.03 | 0.02 |
| | (0.004) | (0.004) | (0.006) |
| ENLACE 9 math subscore | 0.03 | 0.05 | -0.02 |
| | (0.006) | (0.005) | (0.008) |
| ENLACE 9 Spanish subscore | 0.02 | 0.04 | -0.01 |
| | (0.005) | (0.005) | (0.007) |
| Middle school GPA | 0.03 | 0.01 | 0.02 |
| | (0.004) | (0.004) | (0.006) |
| Middle school graduate | 0.08 | 0.08 | -0.00 |
| | (0.005) | (0.005) | (0.007) |

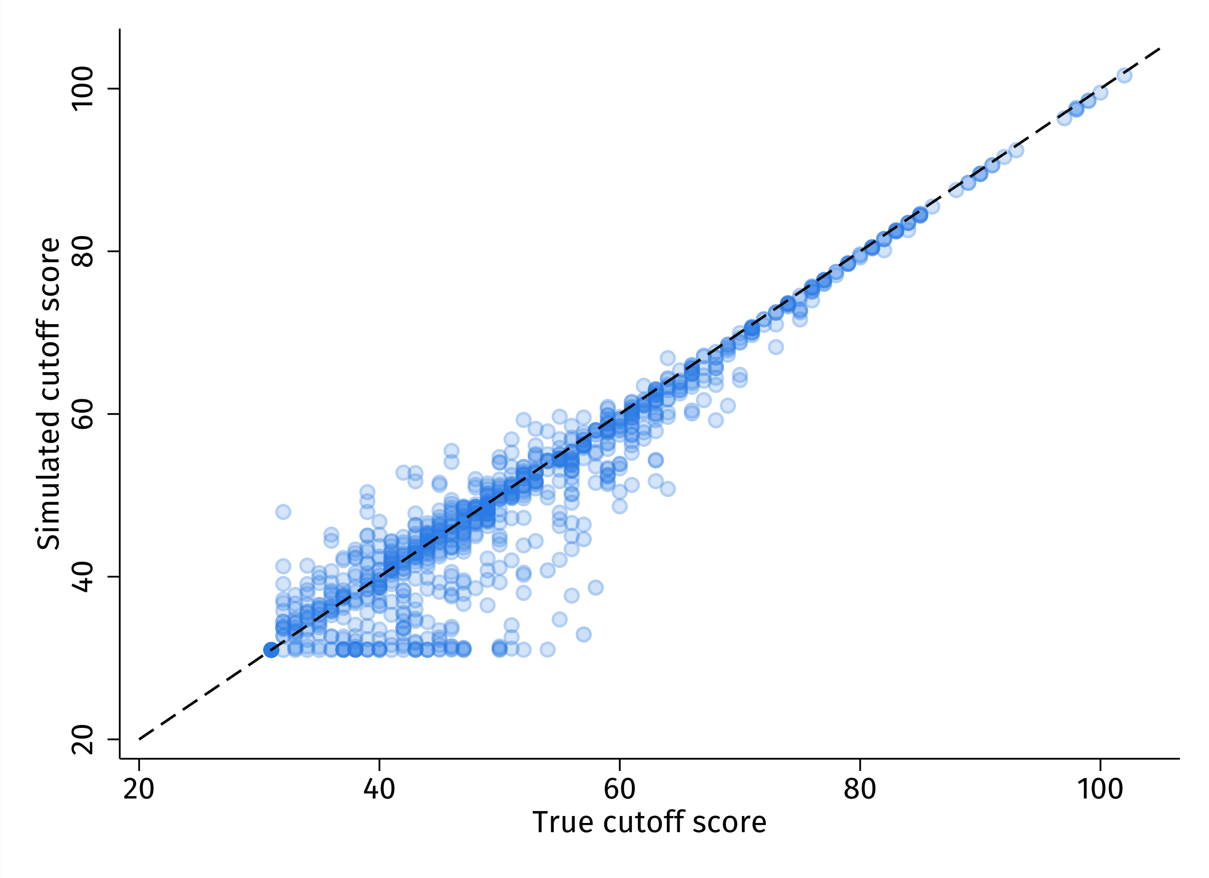
Note: Coefficients in columns 1 and 2 are estimated differences in gender-specific average marginal utilities from the program type indicated in the panel title between students with high and low levels of the indicated characteristic, following Section IV.A. For example, in Panel A, column 1, the "Middle school GPA" entry is the estimated difference in marginal utility from elite STEM programs between males with above-median GPA and males with below-median GPA. Column 3 presents differences between the gender-specific estimates. Standard errors are in parentheses.

## Table C4: School demographics, staffing, and academics by type

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Elite non-STEM | Elite STEM | Traditional | Technical non-STEM | Non-elite STEM |
| **Demographics and staffing** | | | | | |
| Total number of students | 6515.9 | 3279.5 | 1380.4 | 1349.9 | 1360.6 |
| | (3414.00) | (877.54) | (1563.84) | (605.02) | (640.64) |
| Student-to-teacher ratio | 17.6 | 14.5 | 18.3 | 20.3 | 20.0 |
| | (3.06) | (1.75) | (3.69) | (6.69) | (5.60) |
| Percent of teachers that are female | 47.1 | 35.4 | 44.9 | 44.0 | 42.3 |
| | (6.92) | (7.59) | (7.79) | (10.30) | (7.68) |
| Percent of principals that are female | 34.5 | 28.6 | 36.9 | 38.7 | 32.0 |
| | (7.59) | (11.34) | (24.67) | (24.53) | (23.48) |
| Percent of other administrative staff that are female | 53.3 | 50.0 | 58.7 | 50.2 | 53.5 |
| | (4.47) | (3.99) | (12.78) | (9.92) | (10.13) |
| Percent of entering class that is female | 50.9 | 30.0 | 54.3 | 49.6 | 46.5 |
| | (2.59) | (13.16) | (5.62) | (17.32) | (12.06) |
| **Academics** | | | | | |
| Graduation rate | 69.2 | 59.9 | 56.7 | 44.6 | 46.5 |
| | (11.40) | (9.99) | (11.42) | (10.82) | (8.67) |
| Male graduation rate | 61.8 | 57.0 | 48.6 | 39.6 | 40.9 |
| | (10.33) | (9.29) | (12.14) | (10.39) | (7.97) |
| Female graduation rate | 76.5 | 68.1 | 63.8 | 50.1 | 53.6 |
| | (13.02) | (12.11) | (10.80) | (12.04) | (9.56) |
| Failure rate | 7.6 | 43.6 | 31.7 | 29.9 | 31.5 |
| | (16.26) | (6.16) | (8.45) | (8.21) | (9.62) |
| Male failure rate | 8.2 | 44.8 | 36.7 | 32.9 | 34.4 |
| | (17.39) | (5.40) | (6.71) | (8.97) | (9.93) |
| Female failure rate | 7.2 | 41.6 | 28.3 | 28.2 | 28.6 |
| | (15.43) | (7.19) | (9.87) | (8.58) | (9.55) |
| Observations | 18 | 12 | 133 | 44 | 107 |

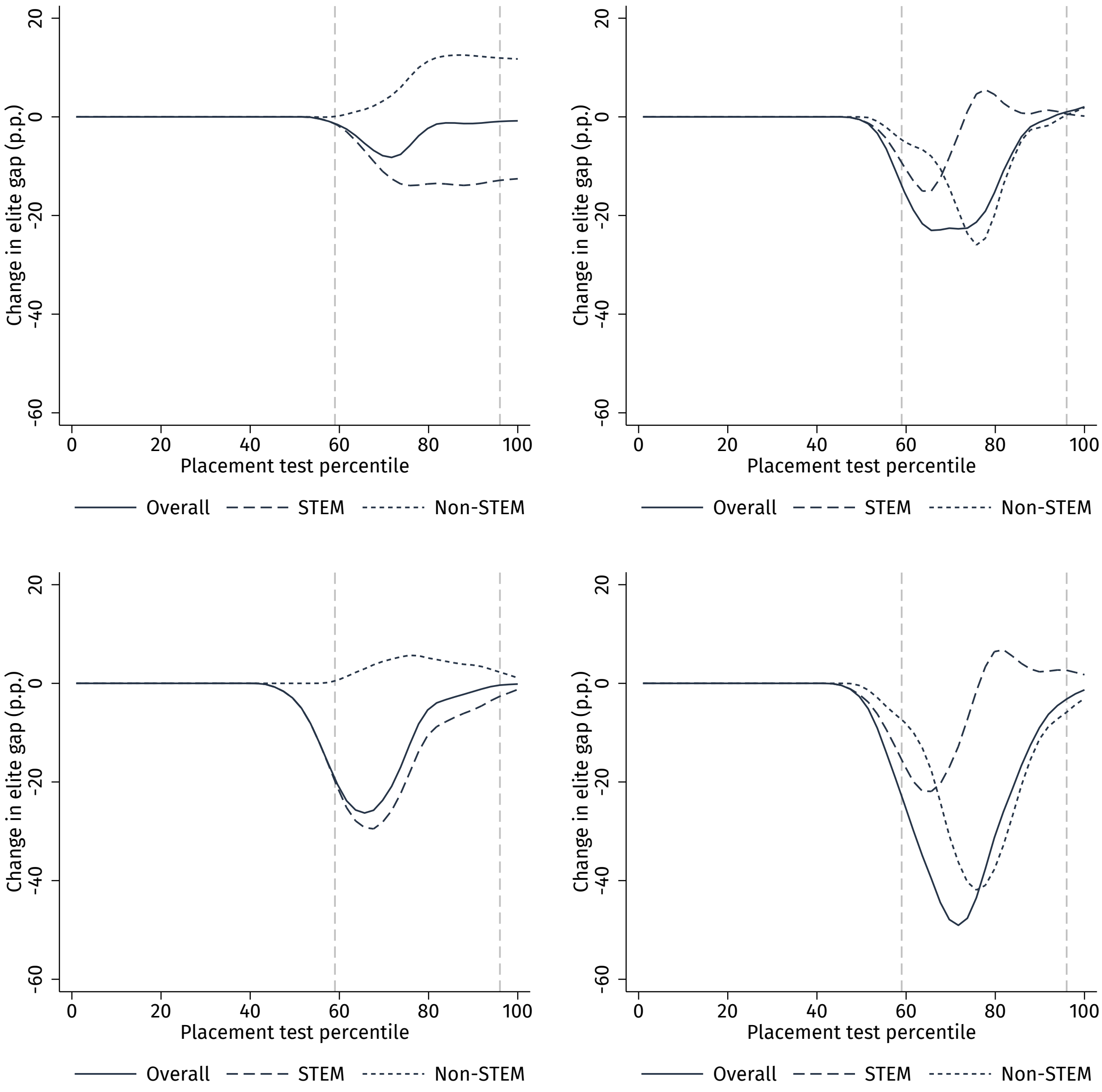
Note: Data are from the school census from 2004 through 2009, which provides data for aggregate campuses as opposed to COMIPEMS programs. Non-elite STEM (technical non-STEM) campuses are classified here as those with 50 percent or more students in STEM (technical non-STEM) programs. Each observation represents the mean values for a campus from 2004 through 2009. Statistics, excluding total number of students, are weighted by average student population of each campus. The graduation rate is computed as the number of graduating students divided by the number of entering students in the corresponding cohort. Failures are defined as students failing between one and five subjects in a given year.

Figure C1: Model fit: simulated versus actual program cutoffs



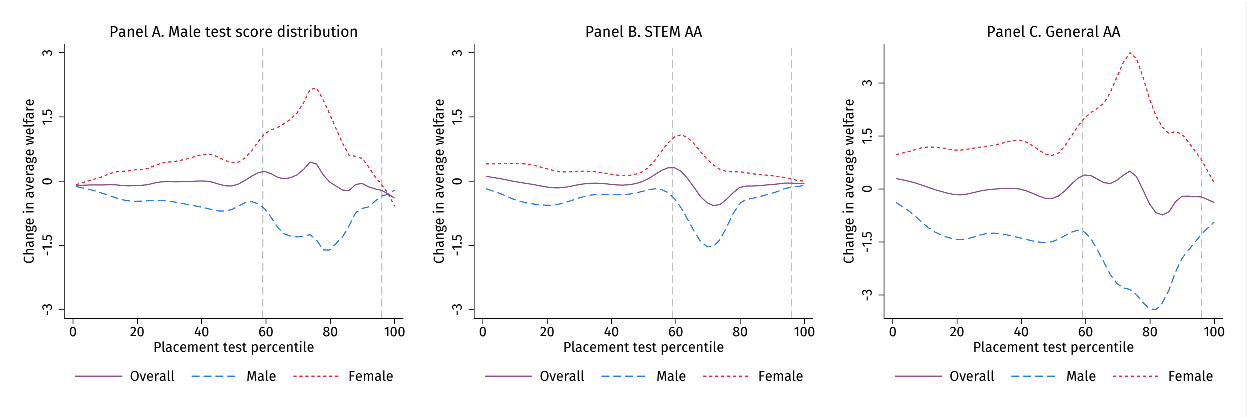
Note: Markers correspond to program-by-year cutoff score pairs, where the x-axis is the true cutoff score and the y-axis is the simulated cutoff score resulting from simulating assignment under the status quo priority structure as described in Section III.C. Simulated cutoff scores are means over 100 independent simulations of the assignment process accounting for uncertainty in student preference parameters, idiosyncratic student preferences, and random tie-breakers in assignment. Cutoff scores are the lowest placement test score a student could obtain and be assigned, and are set to 31 (the minimum to be eligible for assignment) for programs that are not oversubscribed. Opacity is determined by true enrollment counts in the respective year, such that darker points indicate higher enrollment. Dashed line is a 45-degree line. The raw correlation between true and simulated cutoff scores is 0.97 and enrollment-weighted correlation is 0.98.

Figure C2: Simulated effects of preference, score distribution, and priority structure changes on elite gap and its components, by placement test percentile



Note: Lines represent percentage point differences between the simulated elite gaps under the status quo and the counterfactual indicated in the panel title, conditional on the placement test percentile. Simulated changes are means over 100 independent simulations of the assignment process accounting for uncertainty in student preference parameters, idiosyncratic student preferences, and random tie-breakers in assignment. Simulations are as described in Section III.C. Dashed vertical lines indicate the percentiles corresponding to the lowest and highest elite program cutoff scores.

Figure C3: Simulated welfare effects of score distribution and priority structure changes, by gender and placement test percentile



Note: Lines represent, for the indicated subsample, simulated differences in average welfare between the status quo and the counterfactual indicated in the panel title, conditional on the placement test percentile. Simulated changes are means over 100 independent simulations of the assignment process accounting for uncertainty in student preference parameters, idiosyncratic student preferences, and random tie-breakers in assignment. Simulations and welfare computations are as described in Section III.C. Dashed vertical lines indicate the percentiles corresponding to the lowest and highest elite program cutoff scores

# References

**Fairlie, Robert W.** 2017. “Addressing path dependence and incorporating sample weights in the nonlinear Blinder-Oaxaca decomposition technique for logit, probit and other nonlinear models.” *Stanford Institute for Economic Policy Research, Working Paper (17-013).*

**INEGI, Instituto Nacional de Estadística y Geografía.** 2012a. *ENILEMS: Encuesta Nacional de Inserción Laboral de los Egresados de la Educación Media Superior, 2010 and 2012 [dataset].* INEGI. https://www.inegi.org.mx/programas/enilems/2012/ (accessed February 21, 2023).

———. 2012b. *ENOE: National Survey of Occupation and Employment, 2010 and 2012 [dataset].* INEGI. https://www.inegi.org.mx/programas/enoe/15ymas/ (accessed February 21, 2023).

———. 2012c. *National Occupational Classification System (SINCO), 2011. Comparative tables. November 2012 update. [dataset].* INEGI. https://www.inegi.org.mx/contenidos/clasificadoresycatalogos/doc/sinco_tablas_comparativas.xlsx (accessed May 26, 2023).

———. 2017. *EOD: Origin Destination Survey in Households of the Metropolitan Area of the Valley of Mexico, 2017 [dataset].* INEGI. https://www.inegi.org.mx/programas/eod/2017/ http://giitral.iingen.unam.mx/Estudios/EOD-Hogares-01.html (accessed April 5, 2023).

**Ngo, Diana K.L., and Andrew Dustan.** 2023. *Replication data and code for: Preferences, access, and the STEM gender gap in centralized high school assignment.* American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. DOI forthcoming.

**Pathak, Parag A, and Peng Shi.** 2021. “How well do structural demand models work? Counterfactual predictions in school choice.” *Journal of Econometrics* 222 (1): 161–195.

**Rothwell, Jonathan.** 2013. *The hidden STEM economy.* Metropolitan Policy Program at Brookings.