

Finger, Claudia; Solga, Heike

Article — Published Version

Test Participation or Test Performance: Why Do Men Benefit from Test-Based Admission to Higher Education?

Sociology of Education

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Finger, Claudia; Solga, Heike (2023) : Test Participation or Test Performance: Why Do Men Benefit from Test-Based Admission to Higher Education?, Sociology of Education, ISSN 1939-8573, Sage, London, Vol. 96, Iss. 4, pp. 344-366, <https://doi.org/10.1177/00380407231182682>

This Version is available at:
<https://hdl.handle.net/10419/294785>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Online Supplement to Article:

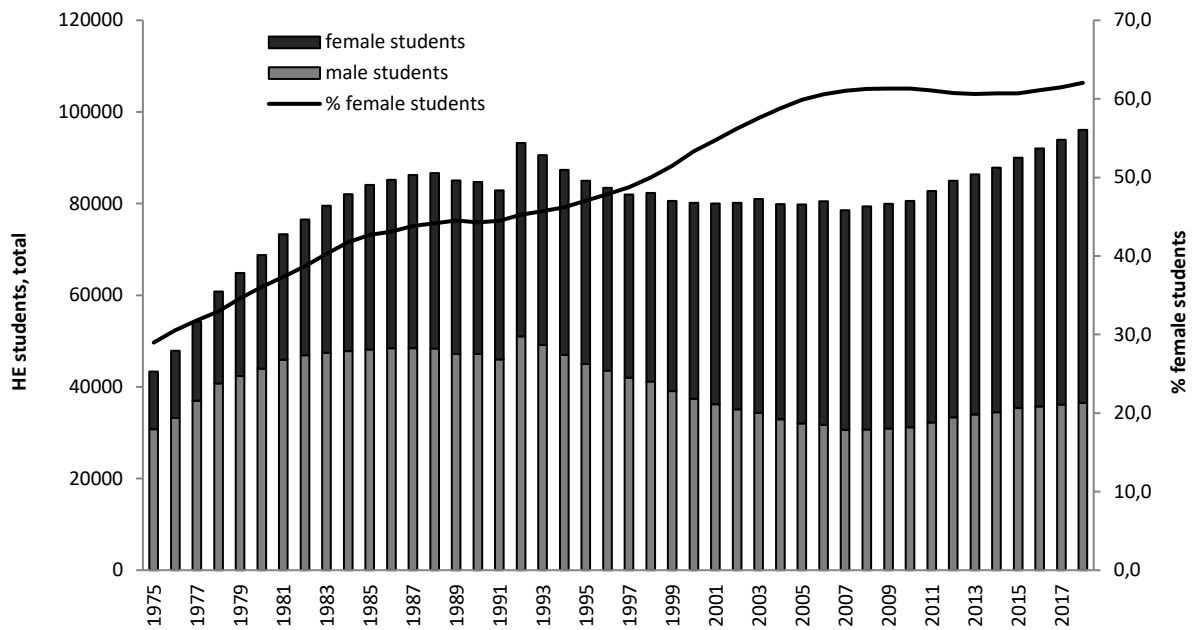
Test participation or test performance: Why do men benefit from test-based admission to higher education?

Table of contents

- A. Development of number of medical school students, by gender
- B. Test-optional versus test-free programs, 2012 to 2018: sources
- C. Sample definition
- D. Descriptive statistics
- E. Test participation: complete models and robustness checks
- F. Test performance: complete models and robustness checks
- G. Admission: complete models
- H. Test participation versus test-score reporting: description of alternative data source
- I. Inclusion of (contextual) social background indicators

Section A. Development of number of medical school students, by gender

Figure A1. Development of the number of male and female students in medical programs in Germany, 1975 to 2018



Note: Until 1989, only former West Germany.

Source: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/lrbil05.html> (retrieved 5-11-2020).

Section B. Test-optional versus test-free programs, 2012 to 2018: sources

Stiftung für Hochschulzulassung: *Studiengänge und Studienorte. Das bundesweite Studienplatzangebot zum Wintersemester [year] an Universitäten:*

- 2012/13; Version: 2012/05/24.
- 2013/14; Version: 2013/04/18.
- 2014/15; Version: 2014/05/13.
- 2015/16; Version: 2015/05/29.
- 2016/17; Version: 2016/06/09.
- 2017/18; Version: 2017/04/27.
- 2018/19; Version: 2018/04/05.

Section C. Sample definition

In this section, we provide further information on sample restrictions and resulting sensitivity analyses to complement the section on “Sample Definition” in the main article.

Due to the different quotas and the stepwise admission procedure (see the Institutional Context section in the main article), it would be neither possible nor meaningful to run our models on the whole population. Different admission rules apply to applicants who already graduated from a higher-education program (2.9 percent) and applicants who were admitted via a special military quota (0.5 percent). These groups are not considered part of our target population and are excluded a priori.

In the following, we provide detailed information on sample restrictions and our study samples for our three analytic steps and suggest some sensitivity analyses to examine potential biases that may result from these sample definitions. As a preparatory step, we listwise deleted some cases with missing information on variables included in the analyses ($N = 145$, 0.05 percent), starting with 293,299 applicants (thereof 35.4 percent men).

For our analyses on test participation, the study sample consists of 223,621 applicants (34.8 percent men, average GPA: 1.95). We first restricted the sample to applicants who (also) applied via the university-admission quota. This means we excluded applicants who only applied via the GPA and/or waiting quota ($N = 28,123$, 9.6 percent of applicants). This step is necessary because tests—our central variable—are only used as part of the university-admission quota (which means information on test participation and performance is only available in this context). This restriction, however, is potentially problematic if only applying via the GPA and waiting quota indicates test-avoidant behavior. Yet as discussed in the main text, the university-admission quota as such is test-optional because test-free programs are available, reporting test scores is not required by test-based programs, and test scores can only improve an applicant’s rank. Thus, it is not necessary to avoid the university-admission quota if one wishes to avoid tests. However, we cannot rule out misperceptions. Therefore, not applying to programs in the university admission quota might be an alternative form of test avoidance. The 10 percent of applicants who did not apply via the university-admission quota have lower GPAs (2.19 versus 1.95 in the study sample), on average, and might therefore be more prone to test anxiety (as theoretically proposed in the Previous Research and Theoretical Considerations section). As men are overrepresented in this group (41 versus 35 percent in the study sample), we might underestimate test avoidance among male applicants and hence

incorrectly specify gender differences in test-taking behaviors. We tested the sensitivity of our results by including applicants who only applied via the GPA and waiting quotas in the sample and categorized them as test-avoiders (see Table E3).

Second, we excluded applicants who applied to a local-test-based program and did not report TMS scores, as they cannot easily be categorized as either test-takers or test-avoiders (see explanation in the Data and Methods section of the main text and Section E of this supplement) ($N = 41,555$, 14.2 percent of applicants). With regard to the share of men (35.5 percent) and average GPA (1.94), they do not differ from cases included in the study sample. As a sensitivity analysis, we re-estimated the models on test participation by categorizing applicants to local-test-based programs (who did not report TMS scores) as test-participants and test-avoiders, respectively, to receive upper- and lower-bound estimates (see Tables E4, and E5, and Figures E2 and E3).

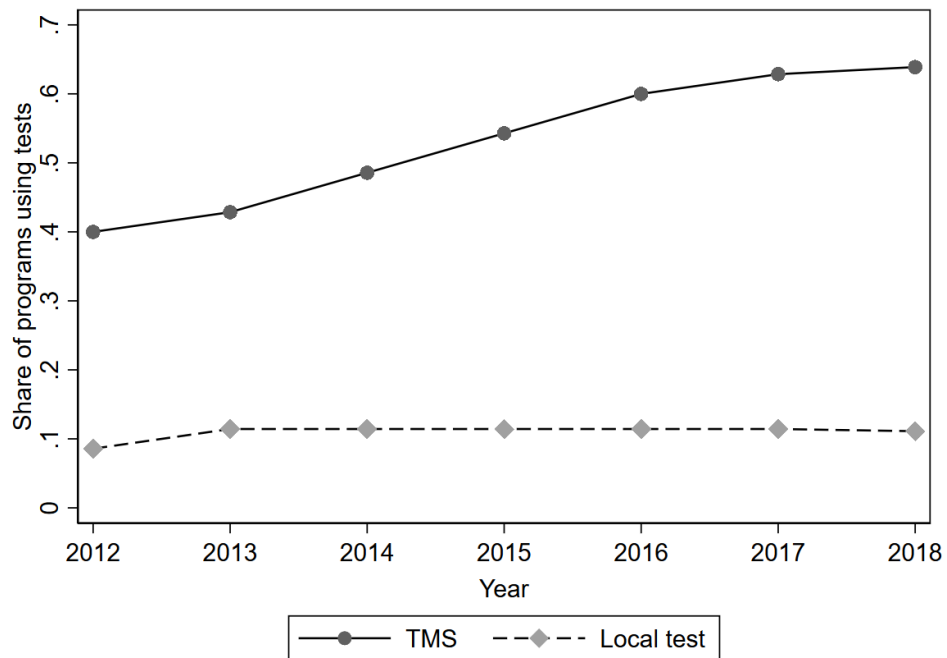
For our analyses on test performance, the study sample consists of 71,187 test-takers (35.2 percent men). Here, we only focused on applicants who participated in the TMS and reported their scores to the clearinghouse, as they are the only cases for which information on test performance exists. This restriction automatically excludes applicants who have already been excluded from the analyses of test participation (see above).

Finally, in our analyses on admission chances, we additionally excluded applicants who were admitted via the GPA or waiting quota (5.4 percent). As each applicant can gain admission to only one program (with a stepwise progression starting with the GPA quota; see Figure 1 in the main text), we (and the applicants) do not know whether they would have been admitted to the programs ranked in the university admission quota. More importantly, for these applicants, test participation and test performance do not have any effect on their admission. Excluding applicants admitted via the GPA and waiting quota does not lead to biased results regarding gender differences in admission chances (overall, 3.0 percent of women and 2.6 percent of men are admitted via the GPA-quota, and 4.2 percent of women and 5.8 percent of men via the waiting quota). This additional last restriction leaves us with 207,872 applicants (34.6 percent men) for analyzing admission chances that are conditional on test participation and with 67,264 applicants (35.1 percent men) for the estimates of admission chances that are conditional on test performance. As a sensitivity analysis, we also provide all results on test participation and test performance with this most restricted sample (see Tables E6 and F2, and Figure E4).

An overview of the different study samples, their gender and GPA composition, and the sensitivity analyses are provided in Table 1 in the main text.

Section D. Descriptive statistics

Figure D1. Share of German medical programs using test scores as a selection criterion, 2012 to 2018



Note: $N = 35$ programs, difference to 100 percent due to programs not using test scores as selection criterion. TMS (*Test für Medizinische Studiengänge* = test for medical programs), local tests are developed and used by three to four universities.

Source: Stiftung für Hochschulzulassung (SfH), annual publications (see Section B of this supplement), authors' calculation.

Table D1. Descriptive statistics, individual-level analyses: test participation

	All <i>N</i> = 223,621			Men <i>N</i> = 77,714			Women <i>N</i> = 145,907		
	%			%			%		
Test participation (scores reported)	32.00			32.42			31.77		
Gender: male	34.75			-			-		
GPA categories									
1.0-1.1	8.77			7.84			9.26		
1.2-1.3	8.81			7.45			9.54		
1.4-1.5	11.55			9.93			12.42		
1.6-1.7	12.71			11.27			13.54		
1.8-1.9	13.09			12.09			13.63		
2.0-2.4	24.13			25.28			23.52		
2.5-2.9	13.98			16.82			12.47		
3.0-4.0	6.94			9.41			5.63		
No. of TMS-based programs applied to (0-6) /TMS scores reported (0/1)									
0/0	7.20			6.94			7.34		
1-6/0	61.14			60.97			61.24		
1/1	2.53			2.44			2.58		
2/1	3.22			3.13			3.27		
3/1	4.52			4.48			4.54		
4/1	6.00			6.02			5.99		
5/1	7.38			7.58			7.27		
6/1	8.01			8.46			7.76		
Control variables									
State of higher-education entrance certificate									
Schleswig-Holstein	2.97			3.03			2.93		
Hamburg	1.63			1.59			1.64		
Lower Saxony	9.11			9.36			8.98		
Bremen	0.93			1.01			0.89		
Northern Rhine-Westfalia	23.35			23.84			23.09		
Hesse	7.82			8.01			7.73		
Rhineland-Palatinate	4.10			4.03			4.14		
Baden-Wuerttemberg	15.75			15.92			15.66		
Bavaria	14.92			14.46			15.16		
Saarland	1.37			1.45			1.32		
Berlin	2.87			2.82			2.89		
Brandenburg	1.78			1.64			1.85		
Mecklenburg-Western Pomerania	1.36			1.35			1.37		
Saxony	3.97			3.88			4.02		
Saxony-Anhalt	1.40			1.31			1.46		
Thuringia	2.12			1.98			2.20		
Abroad	4.56			4.31			4.70		
Year of application									
2012	15.50			16.46			14.99		
2013	14.45			14.90			14.21		
2014	14.04			14.47			13.81		
2015	14.07			14.28			13.96		
2016	14.21			13.82			14.41		
2017	13.82			13.28			14.10		
2018	13.92			12.80			14.52		
	All			Men			Women		
	Min.	Max.	Mean (SD)	Min.	Max.	Mean (SD)	Min.	Max.	Mean (SD)
GPA	1	4	1.95 (0.61)	1	4	2.03 (0.63)	1	4	1.90 (0.59)
Age	17	30	20.84 (2.81)	17	30	21.15 (2.93)	17	30	20.68 (2.73)
No. of applications	1	6	4.94 (1.74)	1	6	4.95 (1.75)	1	6	4.94 (1.74)

Note: German GPA ranges from 1.0 (highest) to 4.0 (lowest). SD = standard deviation.

Sources: SfH application register 2012 to 2018, study sample “test participation” (see the Data and Methods section for definition), authors’ calculation.

Table D2. Descriptive statistics, individual-level analyses: test performance

	All <i>N</i> = 71,187 %			Men <i>N</i> = 25,083 %			Women <i>N</i> = 46,104 %			
Gender: male	35.24			-			-			
GPA categories										
1.0-1.1	7.20			6.30			7.70			
1.2-1.3	8.13			6.82			8.85			
1.4-1.5	14.05			12.22			15.04			
1.6-1.7	17.27			15.57			18.19			
1.8-1.9	16.93			16.14			17.36			
2.0-2.4	24.48			27.35			22.91			
2.5-2.9	9.31			11.95			7.88			
3.0-4.0	2.62			3.64			2.07			
Control variables										
State of higher-education entrance certificate										
Schleswig-Holstein	3.21			3.22			3.20			
Hamburg	2.02			1.82			2.11			
Lower Saxony	8.10			8.28			8.01			
Bremen	0.72			0.89			0.62			
Northern Rhine-Westfalia	19.65			20.35			19.26			
Hesse	7.09			7.41			6.92			
Rhineland-Palatinate	5.37			5.23			5.45			
Baden-Wuerttemberg	21.74			21.01			22.13			
Bavaria	16.91			16.10			17.34			
Saarland	1.24			1.45			1.13			
Berlin	2.39			2.17			2.52			
Brandenburg	1.35			1.28			1.39			
Mecklenburg-Western Pomerania	1.04			1.05			1.04			
Saxony	3.06			3.34			2.90			
Saxony-Anhalt	1.44			1.51			1.40			
Thuringia	1.57			1.62			1.55			
Abroad	3.10			3.24			3.02			
Year of application										
2012	12.62			13.37			12.21			
2013	13.90			14.54			13.55			
2014	14.16			14.89			13.76			
2015	14.66			14.96			14.50			
2016	14.90			14.75			14.98			
2017	14.88			14.16			15.27			
2018	14.89			13.33			15.73			
		All			Men			Women		
	Min.	Max.	Mean (SD)	Min.	Max.	Mean (SD)	Min.	Max.	Mean (SD)	
GPA	1	4	1.84 (0.50)	1	4	1.91 (0.52)	1	4	1.80 (0.48)	
TMS-score	1	4	1.90 (0.53)	1	4	1.82 (0.52)	1	4	1.95 (0.54)	
Inverted TMS-score	1	4	3.09 (0.53)	1	4	3.18 (0.52)	1	4	3.05 (0.54)	
Age	17	30	21.06 (2.53)	17	30	21.38 (2.65)	17	30	20.89 (2.44)	

Note: German GPA ranges from 1.0 (highest) to 4.0 (lowest). SD = standard deviation.

Sources: SfH application register 2012 to 2018, study sample “test-taker” (see the Data and Methods section for definition), authors’ calculation.

Table D3. Descriptive statistics, individual-level analyses: admissions

Sample: all applicants in study sample (N = 207,872)									
	All			Men			Women		
	Min.	Max.	Percent / Mean (SD)	Min.	Max.	Percent / Mean (SD)	Min.	Max.	Percent / Mean (SD)
Admission			21.52			20.66			21.97
Av. GPA of applicants <i>with</i> admission	1	3	1.32 (0.24)	1	2.8	1.35 (0.26)	1	3	1.30 (0.22)
Av. GPA of applicants <i>without</i> admission	1	4	2.13 (0.53)	1	4	2.22 (0.54)	1	4	2.08 (0.51)
Among admitted:									
Admitted to <i>test-based</i> program			58.49			60.61			57.44
Av. GPA of applicants admitted to <i>test-based</i> program	1	3	1.33 (0.26)	1	2.8	1.37 (0.28)	1	3	1.31 (0.24)
Av. GPA of applicants admitted to <i>non-test-based</i> program	1	2.5	1.29 (0.21)	1	2.3	1.32 (0.22)	1	2.5	1.28 (0.19)
Sample: only test-takers (N = 67,264)									
	All			Men			Women		
	Min.	Max.	Percent / Mean (SD)	Min.	Max.	Percent / Mean (SD)	Min.	Max.	Percent / Mean (SD)
Admission			31.36			31.85			31.10
Av. GPA of applicants <i>with</i> admission	1	2.8	1.42 (0.26)	1	2.8	1.46 (0.28)	1	2.6	1.39 (0.25)
Av. GPA of applicants <i>without</i> admission	1	4	2.03 (0.43)	1	4	2.11 (0.44)	1	4	1.99 (0.41)
Av. TMS-score of applicants <i>with</i> admission	1	3.7	1.52 (0.45)	1	3.4	1.45 (0.43)	1	3.7	1.57 (0.46)
Av. TMS-score of applicants <i>without</i> admission	1	4	2.08 (0.47)	1	4	1.99 (0.47)	1	4	2.13 (0.47)
Among admitted:									
Admitted to <i>test-based</i> program			78.30			79.78			77.48
Av. GPA of applicants admitted to <i>test-based</i> program	1	2.8	1.42 (0.26)	1	2.8	1.46 (0.28)	1	2.6	1.40 (0.25)
Av. GPA of applicants admitted to <i>non-test-based</i> program	1	2.2	1.39 (0.24)	1	2.2	1.44 (0.26)	1	2.2	1.37 (0.23)
Av. TMS-score of applicants admitted to <i>test-based</i> program	1	3.4	1.41 (0.39)	1	3.2	1.34 (0.35)	1	3.4	1.45 (0.40)
Av. TMS-score of applicants admitted to <i>non-test-based</i> program	1	3.7	1.93 (0.43)	1	3.4	1.87 (0.42)	1	3.7	1.96 (0.43)

Note: German GPA ranges from 1.0 (highest) to 4.0 (lowest).

Sources: SfH application register 2012 to 2018, study sample “admission” (see the Data and Methods section for definition), authors’ calculation.

Section E. Test participation: complete models and robustness checks

Table E1. Full Table 2 in the main text (display of all control variables):
Gender and GPA effects on test participation (TMS scores reported)

	M1	M2	M3	M4	M5
	All programs	All programs	0-2 TMS-based programs	3-4 TMS-based programs	5-6 TMS-based programs
Gender: male	0.003 (0.002)	0.020*** (0.002)	-0.001 (0.003)	0.012*** (0.004)	0.043*** (0.004)
GPA categories, ref.: 1.0-1.1 (highest)					
1.2-1.3		0.024*** (0.005)	-0.014* (0.006)	0.009 (0.009)	0.080*** (0.011)
1.4-1.5		0.109*** (0.005)	0.006 (0.006)	0.075*** (0.008)	0.170*** (0.010)
1.6-1.7		0.137*** (0.004)	0.024*** (0.006)	0.076*** (0.008)	0.146*** (0.010)
1.8-1.9		0.102*** (0.004)	0.005 (0.006)	0.014 (0.008)	0.089*** (0.010)
2.0-2.4		-0.007 (0.004)	-0.060*** (0.005)	-0.105*** (0.007)	-0.049*** (0.009)
2.5-2.9		-0.125*** (0.004)	-0.103*** (0.005)	-0.226*** (0.008)	-0.217*** (0.010)
3.0-4.0 (lowest)		-0.218*** (0.004)	-0.149*** (0.005)	-0.297*** (0.008)	-0.362*** (0.011)
Control variables					
Age	0.010*** (0.000)	0.022*** (0.000)	0.012*** (0.000)	0.026*** (0.001)	0.025*** (0.001)
Federal state of high school graduation, ref.: NRW					
Schleswig-Holstein	0.069*** (0.006)	0.065*** (0.006)	0.052*** (0.008)	0.039*** (0.011)	-0.005 (0.012)
Hamburg	0.123*** (0.008)	0.106*** (0.008)	0.069*** (0.012)	0.047*** (0.014)	0.016 (0.014)
Lower Saxony	0.012* (0.004)	0.026*** (0.004)	-0.004 (0.004)	-0.002 (0.007)	-0.008 (0.008)
Bremen	-0.027** (0.010)	-0.031*** (0.009)	0.007 (0.012)	-0.060*** (0.017)	-0.165*** (0.022)
Hesse	0.021*** (0.004)	0.015*** (0.004)	-0.012* (0.005)	-0.027*** (0.007)	-0.085*** (0.008)
Rhineland-Palatinate	0.142*** (0.005)	0.136*** (0.005)	0.063*** (0.007)	0.093*** (0.009)	0.092*** (0.010)
Baden-Wuerttemberg	0.179*** (0.003)	0.176*** (0.003)	0.110*** (0.005)	0.108*** (0.006)	0.028*** (0.006)
Bavaria	0.100*** (0.003)	0.097*** (0.003)	0.033*** (0.005)	-0.022*** (0.006)	-0.057*** (0.006)
Saarland	0.022** (0.008)	0.020* (0.008)	-0.003 (0.010)	0.003 (0.014)	-0.016 (0.019)
Berlin	0.006 (0.006)	-0.005 (0.006)	-0.008 (0.007)	-0.032** (0.011)	-0.071*** (0.013)
Brandenburg	-0.022** (0.007)	-0.045*** (0.007)	-0.014* (0.007)	-0.059*** (0.013)	-0.103*** (0.018)

Table continues next page.

Continued Table E1. Full Table 2 in the main text (display of all control variables)

	M1	M2	M3	M4	M5
	All programs	All programs	0-2 TMS-based programs	3-4 TMS-based programs	5-6 TMS-based programs
Mecklenburg-Western Pomerania	-0.025** (0.008)	-0.049*** (0.007)	-0.019* (0.008)	-0.024 (0.015)	-0.072** (0.022)
Saxony	-0.024*** (0.005)	-0.036*** (0.005)	-0.022*** (0.005)	-0.057*** (0.008)	-0.102*** (0.012)
Saxony-Anhalt	0.060*** (0.009)	0.046*** (0.008)	0.062*** (0.011)	0.025 (0.015)	-0.056** (0.018)
Thuringia	-0.033*** (0.006)	-0.060*** (0.006)	-0.039*** (0.007)	-0.102*** (0.010)	-0.189*** (0.014)
Abroad	-0.047*** (0.005)	-0.058*** (0.004)	-0.038*** (0.005)	-0.106*** (0.007)	-0.195*** (0.011)
Year of application, ref. 2012					
2013	0.052*** (0.003)	0.048*** (0.003)	0.037*** (0.004)	0.013 (0.007)	-0.052*** (0.009)
2014	0.067*** (0.003)	0.059*** (0.003)	0.022*** (0.004)	-0.009 (0.006)	-0.077*** (0.009)
2015	0.077*** (0.004)	0.068*** (0.003)	0.021*** (0.004)	-0.016* (0.006)	-0.101*** (0.009)
2016	0.078*** (0.003)	0.065*** (0.003)	0.005 (0.004)	-0.082*** (0.006)	-0.186*** (0.008)
2017	0.087*** (0.004)	0.073*** (0.003)	0.008 (0.005)	-0.082*** (0.006)	-0.188*** (0.008)
2018	0.082*** (0.004)	0.069*** (0.003)	0.008 (0.005)	-0.077*** (0.006)	-0.212*** (0.008)
Number of observations	223,621	223,621	87,369	71,988	64,264
AIC	273,567.5	262,113.1	72,540.0	84,783.9	81,342.6
BIC	273,825.4	262,443.2	72,840.1	85,077.8	81,632.9
R ² (Maddala)	0.030	0.079	0.035	0.083	0.110

Note: Average marginal effects based on logistic regressions, standard errors are in parentheses.

Sources: SfH application register 2012 to 2018, study sample “test participation” (see the Data and Methods section for definition), SfH annual publications (see Section B), authors’ calculation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

Table E2. Gender and GPA effects on test participation (TMS scores reported), including interaction terms

M2 of Table 2 in the main text and M2 to M5 of Table E1, including interaction terms between gender and GPA categories. Figure 2 in the main text is based on these models.

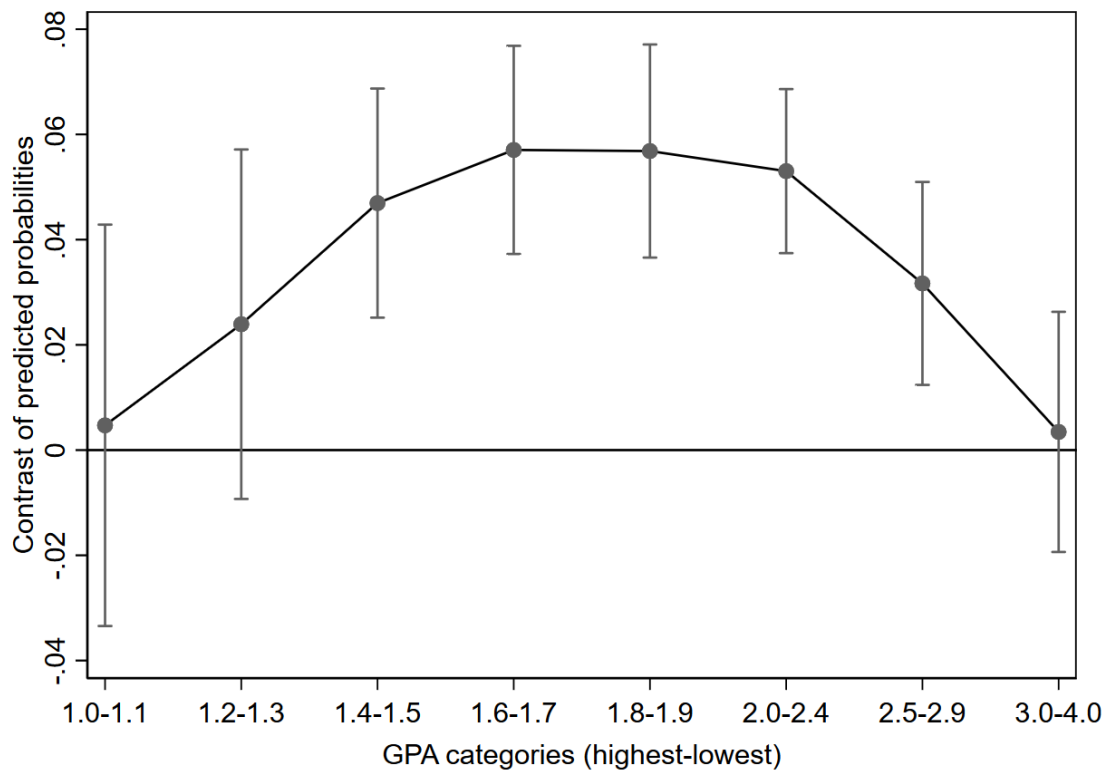
	M2 All programs	M3 0-2 TMS-based programs	M4 3-4 TMS-based programs	M5 5-6 TMS-based programs
Gender: male	0.913* (0.033)	0.946 (0.055)	0.886* (0.055)	1.019 (0.080)
GPA categories, ref. 1.0-1.1 (highest)				
1.2-1.3	1.092 (0.043)	0.903*** (0.043)	1.013 (0.047)	1.387*** (0.080)
1.4-1.5	1.562*** (0.046)	1.012 (0.046)	1.337*** (0.058)	2.052*** (0.107)
1.6-1.7	1.748*** (0.051)	1.039** (0.051)	1.338*** (0.058)	1.801*** (0.091)
1.8-1.9	1.487*** (0.048)	1.046 (0.048)	0.984 (0.043)	1.381*** (0.069)
2.0-2.4	0.884*** (0.027)	0.622*** (0.027)	0.556*** (0.023)	0.766*** (0.037)
2.5-2.9	0.450*** (0.013)	0.379*** (0.020)	0.270*** (0.014)	0.372*** (0.020)
3.0-4.0 (lowest)	0.208*** (0.009)	0.203*** (0.016)	0.141*** (0.010)	0.173*** (0.013)
Gender × GPA categories				
Male×1.2-1.3	1.078 (0.054)	1.024 (0.087)	1.086 (0.093)	1.088 (0.117)
Male×1.4-1.5	1.173*** (0.054)	1.076 (0.086)	1.114 (0.089)	1.246* (0.121)
Male×1.6-1.7	1.212*** (0.054)	1.061 (0.083)	1.127 (0.088)	1.294** (0.120)
Male×1.8-1.9	1.238*** (0.055)	0.947 (0.076)	1.294*** (0.099)	1.259* (0.115)
Male×2.0-2.4	1.307*** (0.053)	1.066 (0.078)	1.305*** (0.092)	1.213* (0.103)
Male×2.5-2.9	1.286*** (0.059)	1.242** (0.104)	1.261** (0.101)	1.135 (0.103)
Male×3.0-4.0	1.169* (0.072)	0.955 (0.112)	1.217 (0.128)	1.006 (0.114)
Control variables	yes	yes	yes	yes
Constant	0.026*** (0.001)	0.025*** (0.002)	0.052*** (0.004)	0.237*** (0.022)
Number of observations	223,621	87,369	71,988	64,264
AIC	262,065.2	72,540.8	84,772.8	81,339.6
BIC	262,467.6	72,906.6	85,131.0	81,693.4
R ² (Maddala)	0.079	0.034	0.084	0.110

Note: Odds ratios, standard errors are in parentheses. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, study sample “test participation” (see the Data and Methods section for definition), SfH annual publications (see Section B), authors’ calculation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

Figure E1. Significance test for difference in predicted probabilities of test participation between male and female applicants, 5 to 6 applications to TMS-based programs



Note: Contrast of predicted probabilities as shown in Figure 2 in the main text (third panel) based on full logistic regression models with interaction terms Gender×GPA categories (see Table E2). Confidence intervals based on two-tailed test. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, study sample “test participation” (see the Data and Methods section for definition), authors’ calculation.

Robustness check 1

Alternative sample: Including applicants who only applied via GPA and/or waiting quota; hypothetical assignment with test participation = 0

Table E3. Gender and GPA effects on test participation (TMS scores reported), alternative sample 1

	M1	M2
Gender: male	-0.006** (0.002)	0.014*** (0.002)
GPA categories, ref.: 1.0-1.1 (highest)		
1.2-1.3		0.017*** (0.005)
1.4-1.5		0.096*** (0.004)
1.6-1.7		0.123*** (0.004)
1.8-1.9		0.086*** (0.004)
2.0-2.4		-0.023*** (0.004)
2.5- 2.9		-0.140*** (0.004)
3.0-4.0 (lowest)		-0.225*** (0.004)
Control variables	yes	yes
Observations	251,744	251,744
AIC	295,578.2	281,475.1
BIC	295,839.1	281,809.1
R^2 (Maddala)	0.020	0.073

Note: Average marginal effects based on logistic regressions, standard errors are in parentheses. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, SfH annual publications (see Section B), authors' calculation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

Robustness check 2

Local test-based programs

Further information and rationale

The TMS (*Test für Medizinische Studiengänge*, or test for medical programs) is the most prominent test format in Germany. Between 2012 and 2018, only three to four of 35 medical programs used other test formats, which we call “local” tests. One prominent example is the HAM-Nat (*Hamburger Naturwissenschaftstest/Hamburg natural science test*), which tests students’ science knowledge at an upper-secondary level in a multiple-choice format (Hissbach, Klusmann, and Hampe 2011¹). Applicants applying to a local-test-based program via the university-admission quota are invited to the university to take the test at a specified date. To receive an invitation, they must rank the university first (thus making it impossible to apply to more than one local-test-based program). Test participation has to be renewed for each application round. According to admission officers for these programs, around 30 percent of those invited to take the test do not participate. This means that not all applicants to local-test-based programs are test-takers. As test-taking is optional, applicants who do not show up for the test are considered, but they are ranked behind those who participated in the test.² Thus, in certain cases, non-participation could be detrimental for applicants, so local tests are not optional to the same degree as the TMS.

Regarding information covered by the register, we only know whether applicants ranked local test-based programs in the university-admission quota. We have no information on either actual test participation or test scores.

Our main analyses thus focus on the TMS, excluding applicants who applied to a local-test-based program and did not report TMS scores, as they cannot easily be categorized as either test-takers or test-avoiders. Our sensitivity analyses (presented below) reinclude these applicants: hypothetical assignment with test participation = 0 (lower bound) and test participation = 1 (upper bound).

¹ Hissbach, Johanna C., Dietrich Klusmann, and Wolfgang Hampe. 2011. “Reliability of a Science Admission Test (HAM-Nat) at Hamburg Medical School.” *GMS Zeitschrift für Medizinische Ausbildung* 28(3): Doc44.

² Information on the share of no-shows and the ranking logic is not publicly available. We contacted admission officers to retrieve this information.

Table E4. Gender and GPA effects on test participation (TMS scores reported), alternative sample 2 (lower-bound estimates)

	M1 All programs	M2 All programs	M3 0-2 TMS- based programs	M4 3-4 TMS- based programs	M5 5-6 TMS- based programs
Gender: male	0.000 (0.002)	0.015*** (0.002)	-0.003 (0.002)	0.008* (0.003)	0.039*** (0.004)
GPA categories, ref.: 1.0-1.1 (highest)					
1.2-1.3		0.024*** (0.004)	-0.009 (0.005)	0.010 (0.008)	0.081*** (0.011)
1.4-1.5		0.094*** (0.004)	0.004 (0.005)	0.055*** (0.008)	0.159*** (0.010)
1.6-1.7		0.113*** (0.004)	0.015** (0.005)	0.042*** (0.008)	0.122*** (0.010)
1.8-1.9		0.082*** (0.004)	-0.002 (0.005)	-0.012 (0.007)	0.069*** (0.010)
2.0-2.4		-0.001 (0.004)	-0.047*** (0.004)	-0.104*** (0.007)	-0.056*** (0.009)
2.5- 2.9		-0.107*** (0.004)	-0.082*** (0.004)	-0.210*** (0.007)	-0.222*** (0.010)
3.0-4.0 (lowest)		-0.187*** (0.004)	-0.117*** (0.004)	-0.272*** (0.007)	-0.362*** (0.010)
Control variables	yes	yes	yes	yes	yes
Observations	265,176	265,176	113,104	84,693	67,379
AIC	299,396.1	288,236.6	79,368.2	93,034.3	85,576.4
BIC	299,658.3	288,572.2	79,676.5	93,333.4	85,868.2
R ² (Maddala)	0.037	0.076	0.034	0.081	0.11

Note: Average marginal effects based on logistic regressions, standard errors are in parentheses. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, SfH annual publications (see Section B), authors' calculation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

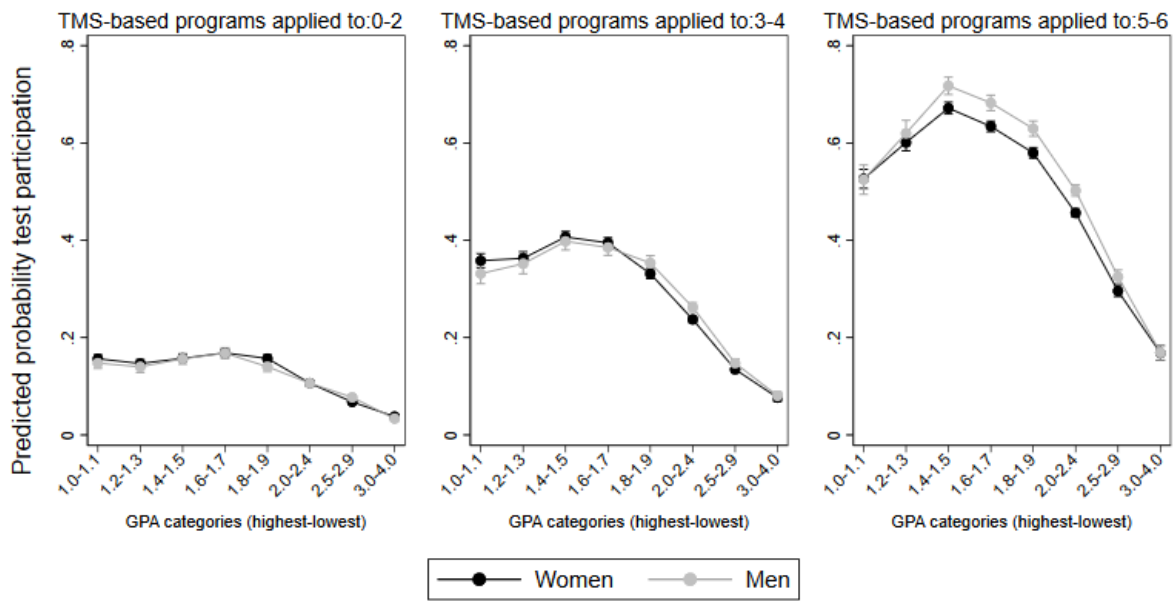
Table E5. Gender and GPA effects on test participation (TMS scores reported), alternative sample 2 (upper-bound estimates)

	M1 All programs	M2 All programs	M3 0-2 TMS- based programs	M4 3-4 TMS- based programs	M5 5-6 TMS- based programs
Gender: male	0.008*** (0.002)	0.021*** (0.002)	0.010*** (0.003)	0.013*** (0.003)	0.043*** (0.004)
GPA categories, ref.: 1.0-1.1 (highest)					
1.2-1.3		0.006 (0.005)	-0.025*** (0.006)	0.008 (0.008)	0.078*** (0.011)
1.4-1.5		0.082*** (0.004)	0.002 (0.006)	0.089*** (0.008)	0.174*** (0.010)
1.6-1.7		0.116*** (0.004)	0.030*** (0.006)	0.115*** (0.008)	0.161*** (0.010)
1.8-1.9		0.088*** (0.004)	0.023*** (0.006)	0.068*** (0.008)	0.106*** (0.010)
2.0-2.4		-0.016*** (0.004)	-0.049*** (0.005)	-0.050*** (0.007)	-0.030** (0.009)
2.5- 2.9		-0.104*** (0.004)	-0.071*** (0.006)	-0.142*** (0.008)	-0.182*** (0.010)
3.0-4.0 (lowest)		-0.173*** (0.005)	-0.109*** (0.006)	-0.200*** (0.008)	-0.313*** (0.011)
Control variables	yes	yes	yes	yes	yes
Observations	265,176	265,176	113,104	84,693	67,379
AIC	351,483.3	343,936.8	135,497.0	110,630.2	85,958.2
BIC	351,745.5	344,272.4	135,805.4	110,929.3	86,250.0
R ² (Maddala)	0.039	0.066	0.091	0.058	0.094

Note: Average marginal effects based on logistic regressions, standard errors are in parentheses. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, SfH annual publications (see Section B), authors' calculation.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

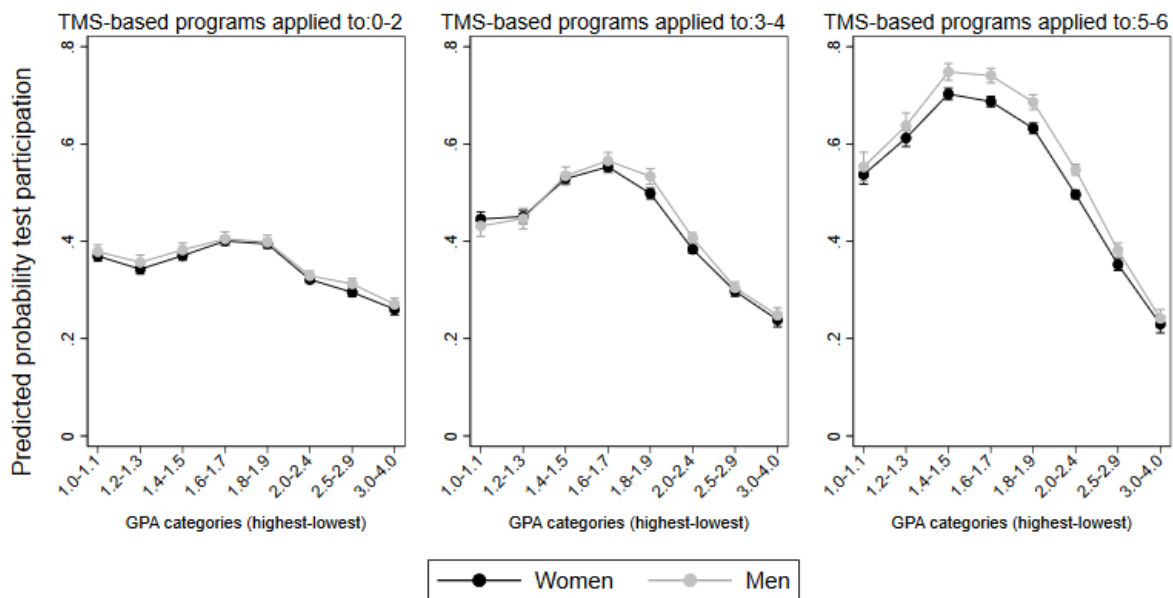
Figure E2. Predicted probabilities of test participation, by gender and GPA categories (95% CI), alternative sample 2 (lower-bound estimates)



Note: Predicted probabilities based on full logistic regression models with interaction terms Gender×GPA categories. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, SfH annual publications (see Section B), authors' calculations.

Figure E3. Predicted probabilities of test participation, by gender and GPA categories (95% CI), alternative sample 2 (upper-bound estimates)



Note: Predicted probabilities based on full logistic regression models with interaction terms gender×GPA categories. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, SfH annual publications (see Section B), authors' calculations.

Robustness check 3

Alternative sample: Excluding applicants with an admission approval in the GPA or waiting quota (“admissions sample”)

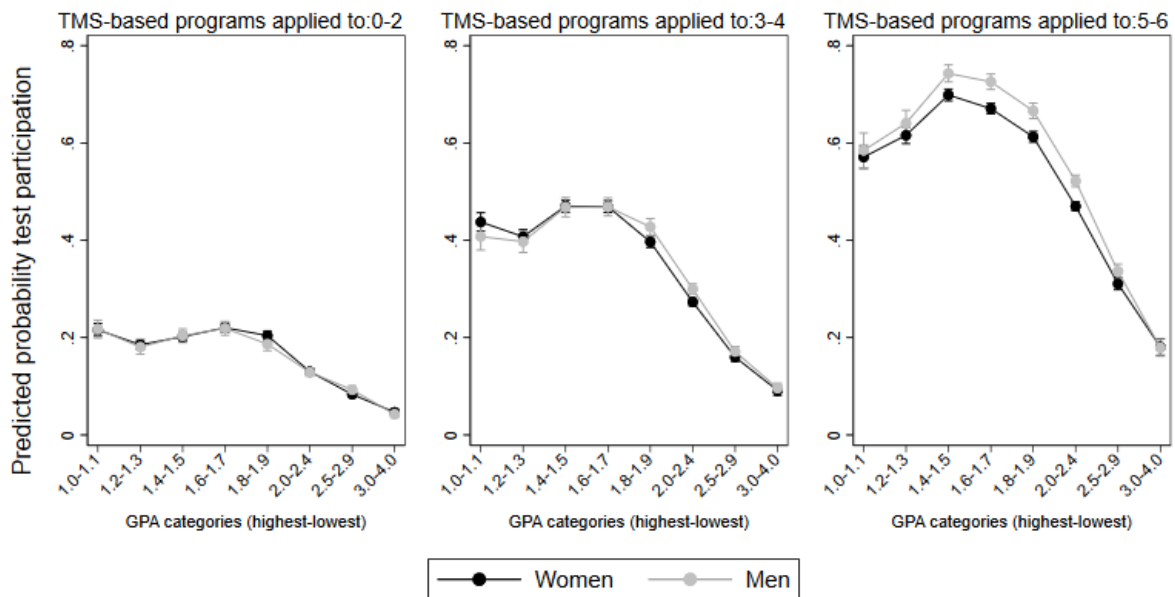
Table E6. Gender and GPA effects on test participation (TMS scores reported), alternative sample 3

	M1 All programs	M2 All programs	M3 0-2 TMS- based programs	M4 3-4 TMS- based programs	M5 5-6 TMS- based programs
Gender: male	0.002 (0.002)	0.020*** (0.002)	-0.002 (0.003)	0.011** (0.004)	0.043*** (0.004)
GPA categories, ref.: 1.0-1.1 (highest)					
1.2-1.3		-0.004 (0.005)	-0.033*** (0.007)	-0.024* (0.010)	0.048*** (0.012)
1.4-1.5		0.080*** (0.005)	-0.014* (0.007)	0.042*** (0.010)	0.136*** (0.011)
1.6-1.7		0.107*** (0.005)	0.003 (0.007)	0.041*** (0.010)	0.111*** (0.011)
1.8-1.9		0.069*** (0.005)	-0.018** (0.007)	-0.022* (0.010)	0.053*** (0.011)
2.0-2.4		-0.044*** (0.005)	-0.087*** (0.006)	-0.146*** (0.009)	-0.089*** (0.011)
2.5- 2.9		-0.163*** (0.005)	-0.129*** (0.006)	-0.264*** (0.009)	-0.258*** (0.011)
3.0-4.0 (lowest)		-0.254*** (0.005)	-0.171*** (0.006)	-0.336*** (0.009)	-0.401*** (0.012)
Control variables	yes	yes	yes	yes	yes
Observations	207,872	207,872	79,306	67,561	61,005
AIC	255,457.6	244,160.4	65,602.6	79,392.7	76,854.1
BIC	255,713.7	244,488.2	65,899.6	79,684.5	77,142.7
R ² (Maddala)	0.032	0.084	0.040	0.090	0.114

Note: Average marginal effects based on logistic regressions, standard errors are in parentheses. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, SfH annual publications (see Section B), authors’ calculation.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

Figure E4. Predicted probabilities of test participation, by gender and GPA categories (95% CI), alternative sample 3



Note: Predicted probabilities based on full logistic regression models with interaction terms Gender×GPA categories. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, SfH annual publications (see Section B), authors' calculations.

Section F. Test performance: complete models and robustness checks

Table F1. Table 3 in the main text, displaying all control variables:
Gender and GPA effects on test performance (only test-takers)

	M1	M2	M3
Gender: male	0.163*** (0.004)	0.183*** (0.004)	0.116*** (0.015)
GPA categories, ref.: 1.0-1.1 (highest)			
1.2-1.3		-0.108*** (0.009)	-0.116*** (0.011)
1.4-1.5		-0.205*** (0.008)	-0.212*** (0.010)
1.6-1.7		-0.266*** (0.008)	-0.279*** (0.010)
1.8-1.9		-0.331*** (0.008)	-0.347*** (0.010)
2.0-2.4		-0.447*** (0.008)	-0.480*** (0.010)
2.5-2.9		-0.644*** (0.010)	-0.702*** (0.012)
3.0-4.0 (lowest)		-0.872*** (0.014)	-0.948*** (0.018)
Gender×GPA-categories			
Male×1.2-1.3			0.024 (0.020)
Male×1.4-1.5			0.022 (0.018)
Male×1.6-1.7			0.043* (0.017)
Male×1.8-1.9			0.053** (0.017)
Male×2.0-2.4			0.100*** (0.016)
Male×2.5-2.9			0.150*** (0.019)
Male× 3.0-4.0			0.182*** (0.027)
Control variables			
Age	-0.050*** (0.001)	-0.013*** (0.001)	-0.013*** (0.001)
Federal state of high school graduation, ref.: NRW			
Schleswig-Holstein	0.116*** (0.011)	0.102*** (0.011)	0.102*** (0.011)
Hamburg	0.051*** (0.014)	0.041** (0.013)	0.039** (0.013)
Lower Saxony	0.083*** (0.008)	0.122*** (0.008)	0.123*** (0.008)
Bremen	-0.075** (0.023)	-0.094*** (0.022)	-0.094*** (0.022)
Hesse	0.093*** (0.008)	0.084*** (0.008)	0.084*** (0.008)
Rhineland-Palatinate	0.163*** (0.009)	0.165*** (0.009)	0.164*** (0.009)
Baden-Wuerttemberg	0.098*** (0.006)	0.097*** (0.006)	0.096*** (0.006)
Bavaria	0.233*** (0.006)	0.228*** (0.006)	0.227*** (0.006)
Saarland	0.049** (0.018)	0.057*** (0.017)	0.056*** (0.017)
Berlin	0.047*** (0.013)	0.038** (0.012)	0.038** (0.012)
Brandenburg	-0.150*** (0.017)	-0.192*** (0.016)	-0.193*** (0.016)
Mecklenburg-Western Pomerania	-0.114*** (0.019)	-0.153*** (0.018)	-0.156*** (0.018)
Saxony	0.032** (0.012)	0.031** (0.011)	0.031** (0.011)
Saxony-Anhalt	-0.081*** (0.016)	-0.082*** (0.016)	-0.083*** (0.016)
Thuringia	-0.101*** (0.016)	-0.153*** (0.015)	-0.154*** (0.015)
Abroad	0.030* (0.012)	-0.052*** (0.011)	-0.051*** (0.011)
Year of application, ref. 2012			
2013	0.024*** (0.007)	0.014 (0.007)	0.013 (0.007)
2014	0.063*** (0.007)	0.042*** (0.007)	0.042*** (0.007)
2015	0.076*** (0.007)	0.044*** (0.007)	0.044*** (0.007)
2016	0.107*** (0.007)	0.064*** (0.007)	0.064*** (0.007)
2017	0.121*** (0.007)	0.066*** (0.007)	0.066*** (0.007)
2018	0.134*** (0.007)	0.072*** (0.007)	0.071*** (0.007)
Constant	3.933*** (0.017)	3.499*** (0.018)	3.524*** (0.019)
Number of observations	71,186	71,186	71,186
<i>R</i> ²	0.106	0.192	0.194

Note: Ordinary least square models, standard errors are in parentheses.

Source: SfH application register 2012 to 2018, study sample “test-taker” (see the Data and Method section for definition), authors’ calculation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

Robustness check 3

Alternative sample: Excluding applicants with an admission approval in the GPA or waiting quota (“admissions sample”)

Table F2. Gender and GPA effects on test performance (only test-takers), alternative sample 3

	M1	M2	M3
Gender: male	0.165*** (0.004)	0.185***(0.004)	0.134***(0.017)
GPA categories, ref.: 1.0-1.1 (highest)			
1.2-1.3		-0.107***(0.010)	-0.110***(0.012)
1.4-1.5		-0.204*** (0.009)	-0.206*** (0.011)
1.6-1.7		-0.264*** (0.009)	-0.272*** (0.011)
1.8-1.9		-0.327*** (0.009)	-0.338*** (0.011)
2.0-2.4		-0.444*** (0.009)	-0.473*** (0.011)
2.5-2.9		-0.653*** (0.011)	-0.706*** (0.013)
3.0-4.0 (lowest)		-0.895*** (0.015)	-0.964*** (0.020)
Gender×GPA-categories			
Male×1.2-1.3			0.007 (0.022)
Male×1.4-1.5			0.004 (0.020)
Male×1.6-1.7			0.026 (0.019)
Male×1.8-1.9			0.035 (0.019)
Male×2.0-2.4			0.086*** (0.019)
Male×2.5-2.9			0.137*** (0.021)
Male× 3.0-4.0			0.162***(0.029)
Control variables	yes	yes	yes
Constant	3.932*** (0.019)	3.506*** (0.019)	3.526*** (0.019)
Number of observations	67,264	67,264	67,264
<i>R</i> ²	0.100	0.187	0.189

Note: Ordinary least square models, standard errors are in parentheses. For control variables included, see Table F1.

Source: SfH application register 2012 to 2018, authors’ calculation.

p* < 0.05; *p* < 0.01; ****p* < 0.001 (two-tailed test).

Section G. Admission: complete models

Table G1. Full Table 4 in the main text (display of all control variables):
Effects of gender, test avoidance, and test performance on admission

	Test participation (TMS scores reported)			Test performance		
	M1	M2	M3	M4	M5	M6
Gender: male	0.001 (0.002)	0.028*** (0.001)	0.025*** (0.001)	0.037*** (0.004)	0.063*** (0.003)	0.016*** (0.002)
GPA categories, ref.: 1.0-1.1 (highest)						
1.2-1.3		-0.120*** (0.003)	-0.126*** (0.003)		-0.057*** (0.004)	-0.085*** (0.006)
1.4-1.5		-0.551*** (0.003)	-0.583*** (0.003)		-0.380*** (0.005)	-0.415*** (0.006)
1.6-1.7		-0.791*** (0.003)	-0.815*** (0.002)		-0.656*** (0.005)	-0.654*** (0.005)
1.8-1.9		-0.897*** (0.002)	-0.903*** (0.002)		-0.829*** (0.004)	-0.802*** (0.005)
2.0-3.0 (lowest)		-0.962*** (0.001)	-0.958*** (0.001)		-0.967*** (0.002)	-0.934*** (0.004)
TMS scores reported			0.122*** (0.001)			
Inverted TMS score (z-stand.)						0.155*** (0.001)
Control variables						
Age	-0.039*** (0.000)	0.008*** (0.000)	0.006*** (0.000)	-0.073*** (0.001)	-0.000 (0.001)	0.009*** (0.001)
Federal state of high school graduation, ref.: NRW						
Schleswig- Holstein	-0.040*** (0.005)	0.018*** (0.004)	0.002 (0.004)	0.030** (0.011)	0.025** (0.008)	-0.000 (0.006)
Hamburg	-0.015* (0.007)	0.020*** (0.005)	0.001 (0.005)	0.047*** (0.013)	0.047*** (0.009)	0.028*** (0.008)
Lower Saxony	-0.071*** (0.003)	0.014*** (0.003)	0.012*** (0.002)	-0.027*** (0.007)	0.046*** (0.006)	0.016*** (0.005)
Bremen	-0.003 (0.009)	-0.009 (0.006)	-0.006 (0.006)	0.008 (0.021)	0.002 (0.015)	0.031* (0.014)
Hesse	-0.008* (0.004)	0.002 (0.002)	-0.001 (0.002)	0.026*** (0.007)	0.016** (0.005)	-0.005 (0.005)
Rhineland- Palatinate	0.024*** (0.005)	0.036*** (0.003)	0.014*** (0.003)	0.043*** (0.008)	0.047*** (0.006)	-0.003 (0.005)
Baden- Wuerttemberg	-0.024*** (0.003)	0.010*** (0.002)	-0.017*** (0.002)	0.018*** (0.005)	0.023*** (0.004)	-0.003 (0.003)
Bavaria	-0.027*** (0.003)	0.003 (0.002)	-0.016*** (0.002)	0.056*** (0.006)	0.061*** (0.004)	-0.009** (0.003)
Saarland	0.002 (0.008)	0.016** (0.005)	0.013** (0.005)	-0.002 (0.015)	0.018 (0.012)	0.001 (0.010)
Berlin	-0.056*** (0.005)	-0.013*** (0.004)	-0.016*** (0.004)	-0.006 (0.012)	0.009 (0.008)	0.000 (0.007)
Brandenburg	0.045*** (0.007)	0.007 (0.004)	0.008* (0.004)	0.005 (0.015)	-0.023* (0.010)	0.034*** (0.010)
Mecklenburg-	0.058*** (0.009)	0.031*** (0.005)	0.031*** (0.005)	0.003 (0.017)	-0.002 (0.012)	0.060*** (0.011)
Western Pomerania						
Saxony	0.033*** (0.005)	0.027*** (0.003)	0.029*** (0.003)	0.002 (0.010)	0.012 (0.008)	0.015* (0.007)

Table continues next page.

Continued Table G1. Full Table 4 in the main text (display of all control variables)

	Test participation (TMS scores reported)			Test performance		
	M1	M2	M3	M4	M5	M6
Saxony-Anhalt	0.033*** (0.008)	0.015** (0.005)	0.005 (0.005)	-0.055*** (0.014)	-0.029** (0.010)	0.010 (0.009)
Thuringia	0.110*** (0.007)	0.002 (0.004)	0.010* (0.004)	0.040** (0.014)	-0.024* (0.010)	0.023* (0.009)
Abroad	0.199*** (0.006)	-0.017*** (0.003)	-0.012*** (0.003)	0.119*** (0.011)	-0.008 (0.008)	0.002 (0.007)
Year of application, ref. 2012						
2013	-0.006 (0.003)	-0.020*** (0.002)	-0.028*** (0.002)	-0.018** (0.006)	-0.033*** (0.005)	-0.035*** (0.004)
2014	-0.006 (0.003)	-0.028*** (0.002)	-0.039*** (0.002)	-0.013* (0.006)	-0.041*** (0.005)	-0.051*** (0.004)
2015	-0.009** (0.003)	-0.043*** (0.002)	-0.057*** (0.002)	-0.012* (0.006)	-0.063*** (0.005)	-0.074*** (0.004)
2016	-0.016*** (0.003)	-0.059*** (0.002)	-0.073*** (0.002)	-0.012 (0.006)	-0.074*** (0.005)	-0.090*** (0.004)
2017	-0.007* (0.003)	-0.061*** (0.002)	-0.078*** (0.002)	0.019** (0.006)	-0.070*** (0.005)	-0.088*** (0.004)
2018	-0.006 (0.003)	-0.066*** (0.002)	-0.084*** (0.002)	0.024*** (0.006)	-0.081*** (0.005)	-0.100*** (0.004)
Number of applications	-0.004*** (0.001)	0.025*** (0.000)	0.017*** (0.000)	-0.043*** (0.001)	0.007*** (0.001)	0.014*** (0.001)
Number of observations	207,872	207,872	207,872	67,264	67,264	67,264
AIC	203,765.3	98,904.6	89,758.3	74,142.4	46,337.4	32,582.8
BIC	204,031.7	99,242.7	90,106.6	74,379.4	46,620.1	32,874.5
R ² (Maddala)	0.059	0.432	0.457	0.133	0.427	0.533

Note: Average marginal effects based on logistic regressions, standard errors are in parentheses.

Sources: SfH application register 2012 to 2018, study sample “admission” (see the Data and Method section for definition), SfH annual publications (see Section B), authors’ calculation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

Table G2. Effects of gender, test participation, and test performance on admission, including interaction terms

M3 and M6 of Table 4 in the main text, including interaction terms between gender and GPA categories. Figures 4 and 5 in the main text are based on these models. Additional model with interaction terms between gender and number of test-based programs applied to, only test participants. Figure 4 (right panel) is based on this model.

	Test participation	No. of TMS-based programs applied to	Test performance
	M3 + interaction	Only test-taker	M6 + interaction
Gender: male	1.163*** (0.030)	1.43*** (0.142)	0.129*** (0.035)
GPA categories, ref.: 1.0-1.1 (highest)			
1.2-1.3	0.169*** (0.007)	0.194*** (0.027)	-1.755*** (0.153)
1.4-1.5	0.012*** (0.001)	0.018*** (0.002)	-4.640*** (0.147)
1.6-1.7	0.002*** (0.000)	0.005*** (0.001)	-6.299*** (0.149)
1.8-1.9	0.001*** (0.000)	0.001*** (0.000)	-7.594*** (0.153)
2.0-3.0 (lowest)	0.000*** (0.000)	0.000*** (0.000)	-10.037*** (0.162)
TMS scores reported	5.413*** (0.129)		
Male×TMS scores reported	1.580*** (0.058)		
Inverted TMS score (z-stand.)			2.003*** (0.027)
Male×Inverted TMS score (z-stand.)			0.196*** (0.041)
No. of TMS-based programs applied to (1-6), ref: 1			
2		1.060 (0.085)	
3		1.542*** (0.120)	
4		2.096*** (0.162)	
5		3.744*** (0.290)	
6		10.295*** (0.808)	
Gender×No. of TMS-based programs			
Male×applied to 2		1.207 (0.161)	
Male×3		1.165 (0.143)	
Male×4		1.209 (0.141)	
Male×5		1.201 (0.135)	
Male×6		1.332** (0.146)	
Control variables	yes	yes	yes
Constant	2.396*** (0.244)	4.758*** (0.520)	2.092*** (0.226)
Number of observations	207,872	67,263	67,263
AIC	90,082.1	42,618.66	32,561.4
BIC	90,420.2	42,992.23	32,862.2
R ² (Maddala)	0.456	0.455	0.533

Note: Odds ratios, standard errors are in parentheses. For control variables included, see Table G1.

Sources: SfH application register 2012 to 2018, study sample “admission” (see the Data and Method section for definition), SfH annual publications (see Section B), authors’ calculation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

Section H. Test participation versus test-score reporting: description of alternative data source

Online panel of 2018 applicant cohort

In cooperation with the central clearinghouse (SfH), we invited all medical school applicants for the winter semester of 2018 to participate in the first online survey via email. The first survey was completed by 7,349 applicants (response rate: 17 percent). This number is comparable to the response rate of a representative large-scale online survey among German university students (Becker, Baillet, and Weber 2019:20). In the second wave, 4,619 respondents participated (63 percent of Wave 1 respondents). The data contain information on whether applicants participated in the TMS and, if so, their test scores (but not whether they submitted test scores). The data are available online (Finger, Wetter, and Solga 2023).

For the additional analyses presented in the main text, we defined the samples for test participation and test performance as close as possible to the study samples on which the main analyses are based (see the Data and Method section of the main text for details).

Furthermore, we constructed sampling weights based on the study samples of the 2018 register data to correct for selective survey participation and attrition. Following recommendations on how to create and apply weights in Stata (Dupraz 2013), we constructed weights using distributions of the following variables:

- gender
- age (grouped)
- federal state in which the university entrance certificate was acquired
- GPA
- whether applicants applied via the GPA and waiting quota
- quota-specific admission rates
- number of TMS-based programs applied to
- interaction terms between gender and all further variables (except for age and federal state)

References

Becker, Karsten, Florence Baillet, and Anna Weber. 2019. 21. "Sozialerhebung. Daten- und Methodenbericht." Hannover: FDZ-DZHW.

Dupraz, Yannic. 2013. "Using Weights in Stata." (retrieved May 09, 2022, [http://www.parisschoolofeconomics.eu/docs/dupraz-yannick/using-weights-in-stata\(1\).pdf](http://www.parisschoolofeconomics.eu/docs/dupraz-yannick/using-weights-in-stata(1).pdf)).

Finger, Claudia, Rebecca Wetter, and Heike Solga. 2023. "Zugang zu medizinischen und pharmazeutischen Studiengängen in Deutschland: Bewerber*innenbefragung (Wintersemester 2018/19). Version 1.0.0." Wissenschaftszentrum Berlin für Sozialforschung (WZB) (<https://dx.doi.org/10.7802/2515>).

Section I. Inclusion of (contextual) social background indicators

The register data on applicants to medical schools do not contain information on applicants' social background. However, given the female majority among applicants, higher test participation and better test performance of the much smaller group of male applicants (around 35 percent, see Table D1) could also be caused by a higher socioeconomic selectivity of male applicants. Yet we have information on applicants' residential postal code, which allows us to add contextual, sociodemographic information to the data. Overall, Germany is divided into more than 8,000 postal codes. We use the following socio-economic and demographic indicators at the postal-code level (all z -standardized): share of long-term unemployed, share of three income groups (annual household income, in Euros: >60,000; 30,000 to 60,000; < 30,000), and share of residents with migration backgrounds. Below, we present our main models on test participation, performance, and admission chances including these contextual socio-economic variables. Additionally, we display interaction effects between gender and these variables.

Most importantly, we find the gender effect does not change after inclusion of these variables, suggesting our findings are not confounded by differences in the social selectivity of male and female applicants. Moreover, we rarely find indications that (contextual) socio-economic variables moderate the gender effect (with the exemption of the share of residents with migration backgrounds, which seems to positively affect test participation rates of women only).

Regarding the effect of the contextual variables on our dependent variables, the share of long-term unemployed residents and the share of households in the highest income group slightly influences test participation and performance. However, the effect is not very substantial and smaller than the gender effect (e.g., 1 SD increase in the share of long-term unemployed residents decreases the probability of test participation by 1.3 percentage point). This small effect might not be surprising given that we look at the positively selected group of applicants and include GPA.

Table I.1. Gender and social background effects on test participation (TMS scores reported)

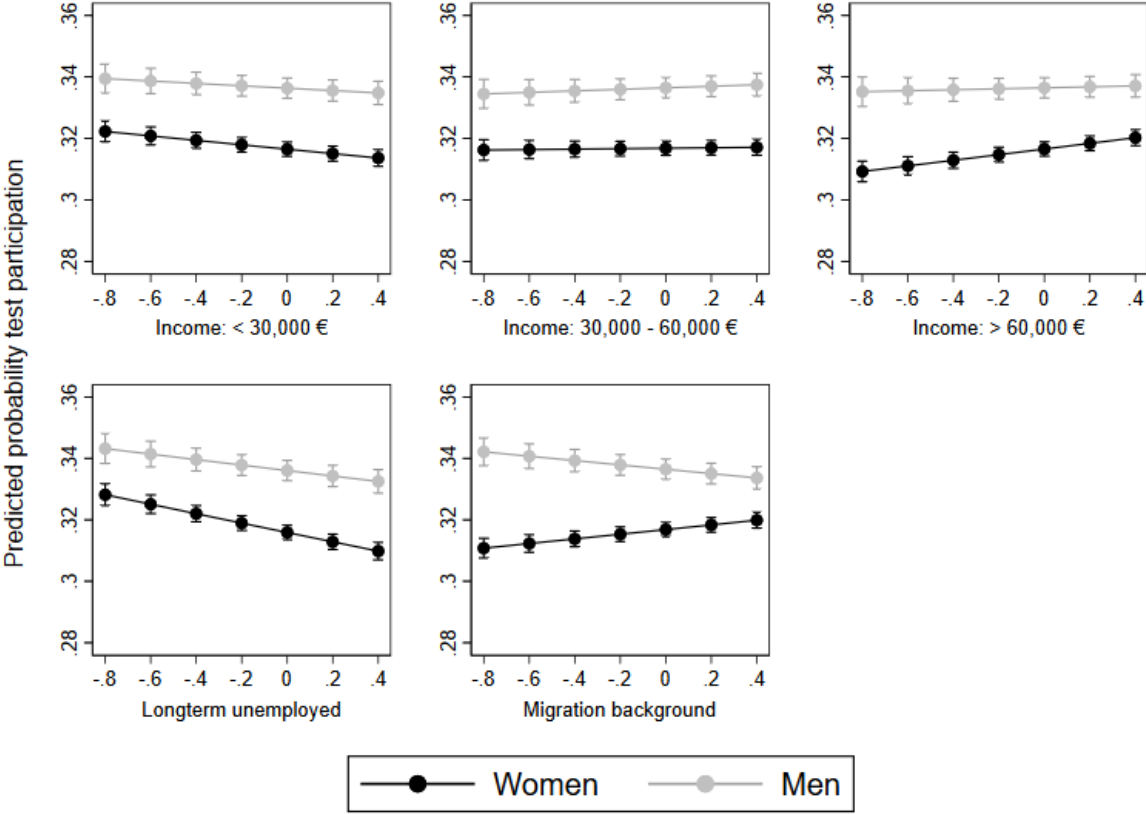
	M2
Gender: Male	0.020*** (0.002)
Social composition of postal code area	
Average annual household income: < 30,000 €	-0.006*** (0.001)
Average annual household income: 30,000 to 60,000 €	0.001 (0.001)
Average annual household income: > 60,000 €	0.007*** (0.001)
Share of long-term unemployed residents	-0.013*** (0.002)
Share of residents with migration background	0.00** (0.001)
GPA categories, ref.: 1.0-1.1 (highest)	
1.2-1.3	0.029*** (0.005)
1.4-1.5	0.113*** (0.005)
1.6-1.7	0.141*** (0.005)
1.8-1.9	0.104*** (0.005)
2.0-2.4	-0.006 (0.004)
2.5- 2.9	-0.127*** (0.004)
3.0-4.0 (lowest)	-0.221*** (0.004)
Control variables	yes
Observations	215,835
AIC	253,579
BIC	253,959.5
<i>R</i> ² (Maddala)	0.081

Note: Average marginal effects based on logistic regressions, standard errors are in parentheses. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, study sample “test participation” (see the Data and Method section for definition; fewer observations due to missing information on applicants’ ZIP code), SfH annual publications (see Section B), authors’ calculation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

Figure I.1. Predicted probabilities of test participation, by gender and social background (95% CI)



Note: Predicted probabilities based on full logistic regression models with interaction terms Gender×Social background variables. For control variables included, see Table E1.
Sources: SfH application register 2012 to 2018, study sample “test participation,” SfH annual publications (see Section B), authors’ calculations.

Table I.2. Gender and social background effects on test performance (test-takers only)

	M3	M3b
Gender: Male	0.183*** (0.004)	0.183*** (0.004)
Social composition of postal code area		
Average annual household income: < 30,000 €	-0.012*** (0.003)	-0.014*** (0.003)
Average annual household income: 30,000 to 60,000 €	0.006* (0.002)	0.004 (0.003)
Average annual household income: > 60,000 €	0.024*** (0.002)	0.023*** (0.003)
Share of long-term unemployed residents	-0.017*** (0.003)	-0.020*** (0.003)
Share of residents with migration background	0.004 (0.002)	0.007** (0.003)
GPA categories, ref.: 1.0-1.1 (highest)		
1.2-1.3	-0.118*** (0.010)	-0.118*** (0.010)
1.4-1.5	-0.214*** (0.009)	-0.214*** (0.009)
1.6-1.7	-0.276*** (0.008)	-0.276*** (0.008)
1.8-1.9	-0.341*** (0.008)	-0.341*** (0.008)
2.0-2.4	-0.455*** (0.008)	-0.455*** (0.008)
2.5-2.9	-0.653*** (0.010)	-0.653*** (0.010)
3.0-4.0 (lowest)	-0.883*** (0.014)	-0.883*** (0.014)
Gender×Social composition		
Male×average annual household income: < 30,000 €		0.005 (0.005)
Male×average annual household income: 30,000 to 60,000 €		0.005 (0.005)
Male×average annual household income: > 60,000 €		0.003 (0.005)
Male×share of long-term unemployed residents		0.007 (0.005)
Male×share of residents with migration background		-0.008 (0.004)
Control variables	yes	yes
Observations	69,491	69,491
Constant	3.504*** (0.019)	3.503*** (0.019)
<i>R</i> ²	0.200	0.200

Note: Ordinary least square models, standard errors are in parentheses. Inverted test scores (1/low – 4/high). For control variables included, see Table F1.

Source: SfH application register 2012 to 2018, study sample “test-taker” (see the Data and Method section for definition; fewer observations due to missing information on applicants’ ZIP code), authors’ calculation.

p* < 0.05; *p* < 0.01; ****p* < 0.001 (two-tailed test).

Table I.3. Gender and social background effects on admission chances

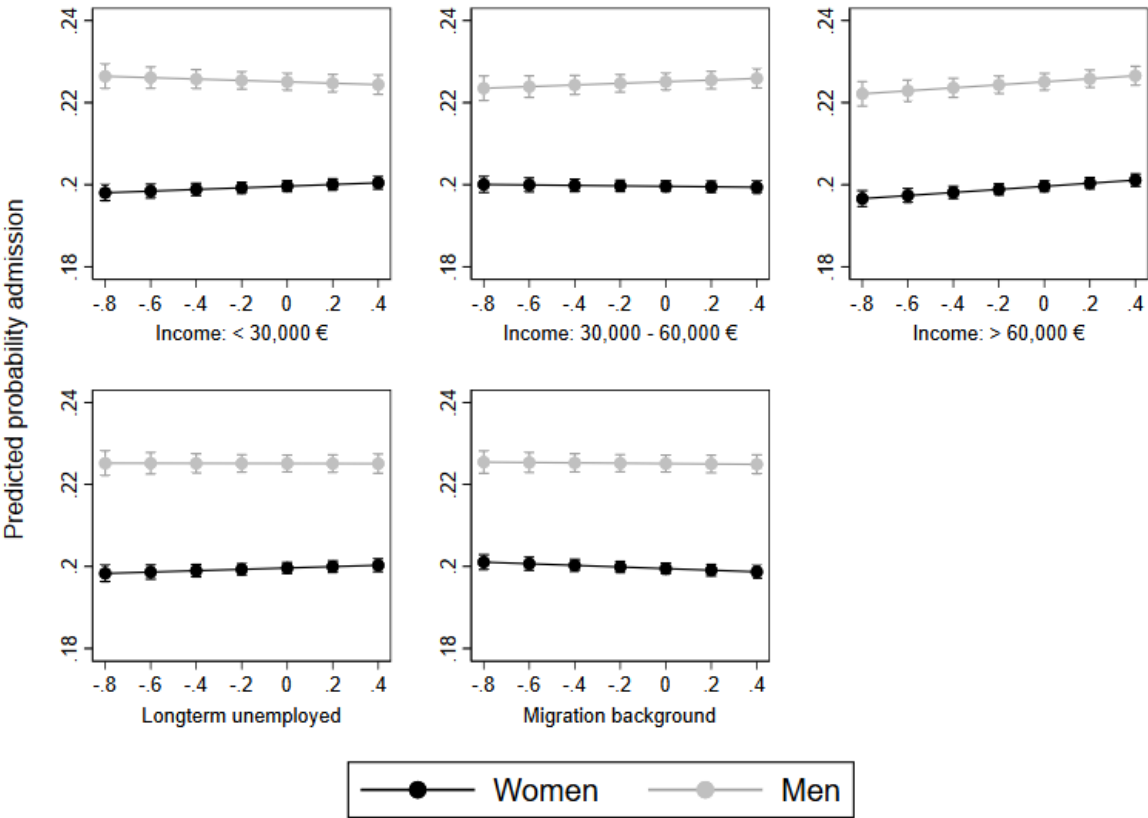
	Test participation (TMS scores reported) M3	Test performance (test-takers only) M6
Gender: Male	0.025*** (0.001)	0.016*** (0.002)
Social composition of postal code area		
Average annual household income: < 30,000 €	0.001 (0.001)	0.002 (0.002)
Average annual household income: 30,000 to 60,000 €	0.000 (0.001)	-0.001 (0.001)
Average annual household income: > 60,000 €	0.004*** (0.001)	0.003* (0.001)
Share of long-term unemployed residents	0.001 (0.001)	0.002 (0.002)
Share of residents with migration background	-0.001* (0.001)	-0.003* (0.001)
GPA categories, ref.: 1.0-1.1 (highest)		
1.2-1.3	-0.121*** (0.003)	-0.078*** (0.006)
1.4-1.5	-0.578*** (0.003)	-0.406*** (0.006)
1.6-1.7	-0.811*** (0.002)	-0.647*** (0.006)
1.8-1.9	-0.900*** (0.002)	-0.795*** (0.006)
2.0-3.0 (lowest)	-0.955*** (0.002)	-0.928*** (0.005)
Test participation	0.122*** (0.001)	
Test scores (z-standardized)		0.156*** (0.001)
Control variables	yes	yes
Observations	201,502	65,799
AIC	867,61.93	318,68.83
BIC	871,39.84	322,05.32
<i>R</i> ² (Maddala)	0.446	0.530

Note: Average marginal effects based on logistic regressions, standard errors are in parentheses. For control variables included, see Table G1.

Sources: SfH application register 2012 to 2018, study sample “admissions” (see Data and Method section for definition; fewer observations due to missing information on applicants’ ZIP code), SfH annual publications (see Section B), authors’ calculation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test).

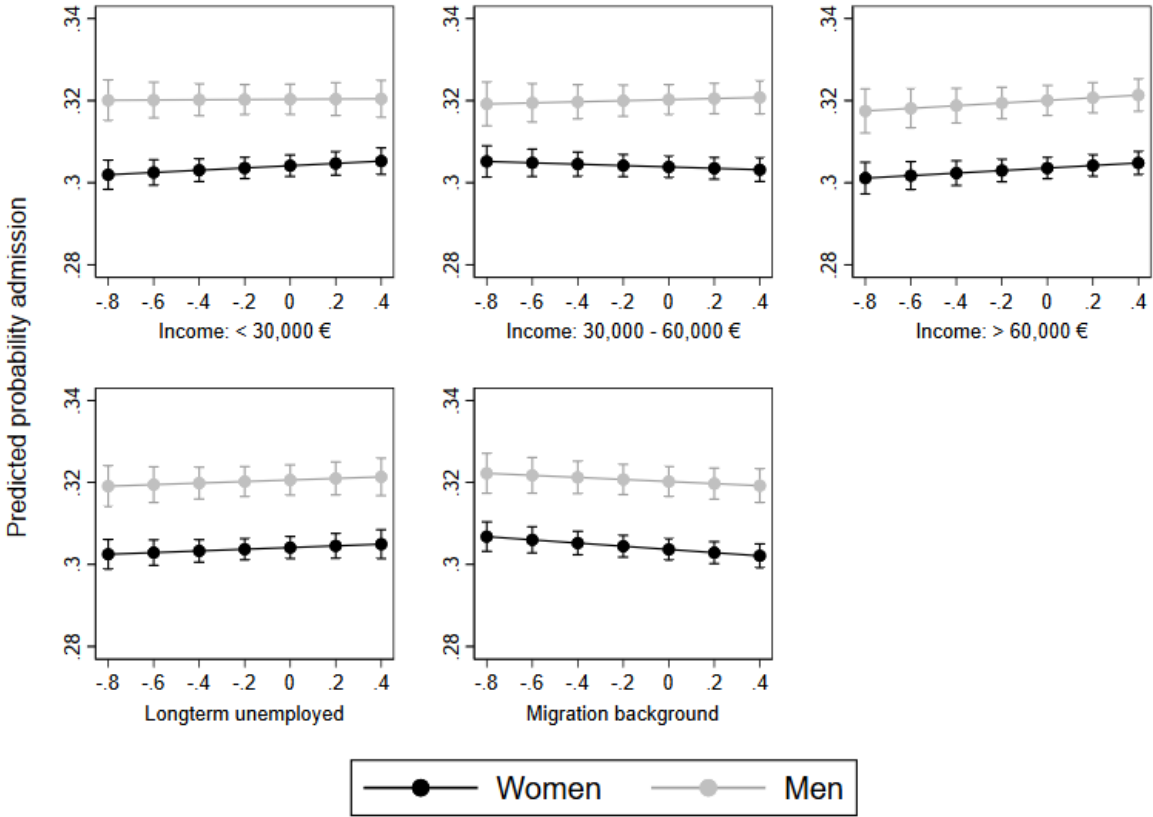
Figure I.2. Predicted probabilities of admission, by gender and social background (95% CI)



Note: Predicted probabilities based on full logistic regression models with interaction terms gender×social background variables. For control variables included, see Table G1.

Sources: SfH application register 2012 to 2018, study sample “admission, conditional on test participation,” SfH annual publications (see Section B), authors’ calculations.

Figure I.3. Predicted probabilities of admission, by gender and social background (95% CI), “test performance” sample



Note: Predicted probabilities based on full logistic regression models with interaction terms gender×social background variables. For control variables included, see Table E1.

Sources: SfH application register 2012 to 2018, study sample “admission, conditional on test performance,” SfH annual publications (see Section B), authors’ calculations.