

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/>

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

Sociology of Education
2023, Vol. 96(4) 344–366
© The Author(s) 2023

DOI: 10.1177/00380407231182682
journals.sagepub.com/home/soe



Claudia Finger¹  and Heike Solga^{1,2}

Abstract

This study illuminates the male advantage in test-based admissions to higher education. In contrast to many other countries, admission tests in Germany are optional, and test-free programs are available. This context offers a unique opportunity to investigate whether the male advantage in test-based admissions is caused by gender differences in test performance or in test participation. We use novel register data for the whole population of 300,000 applicants to highly selective and prestigious medical programs in Germany. We find that men perform better in tests and that female applicants are more likely to withdraw from admission tests. Both differences, however, depend on high school grade point average (GPA): The male advantage in test performance emerges only among test-takers with a lower GPA, and female applicants' stronger test avoidance appears only among women with a medium GPA. Ultimately, both mechanisms contribute to a male advantage in test-based admissions (*ceteris paribus* of GPA), with better test performance being the major source for male applicants' higher admission chances. As a consequence, we find the female advantage in school performance and the male advantage in test-based admissions almost neutralize each other.

Keywords

gender inequality, higher education, admission tests, test participation, test performance, prestigious fields of study, medical schools

In response to the enormous expansion of higher education over the past century, standardized tests are widely used for admissions. These tests promise to help select higher education applicants in an efficient and merit-based manner (Furuta 2017; Grodsky, Warren, and Felts 2008). The “emergence of a test-score meritocracy” (Alon and Tienda 2007:489) for admissions to selective institutions and fields of study has been observed in the United States and numerous other countries (e.g., Australia: Puddey and Mercer 2014; Japan: Koza 2006; Russia: Jackson, Khavenson, and Chirkina 2020; Sweden: Berggren 2007; UK: Tiffin et al. 2014; US: Dunleavy et al. 2013). Yet test-based

admissions have been criticized because the strong reliance on test scores further advantages already advantaged groups, such as high socioeconomic status (SES) or racial majority students (Grodsky et al. 2008; Zwick 2019). A well-established mechanism behind these inequalities are group

¹WZB Berlin Social Science Center, Berlin, Germany

²Freie Universität Berlin, Berlin, Germany

Corresponding Author:

Claudia Finger, WZB Berlin Social Science Center,
Reichpietschufer 50, Berlin, 10785, Germany.
Email: claudia.finger@wzb.eu

differences in resources, which contribute to differential test performance and thus admission chances (Buchmann, Condron, and Roscigno 2010).

Differences in test participation or avoidance can be considered as a second contributor to social inequality in test-based college admissions given that students from socially disadvantaged groups are less likely to take college admission tests (Klaskik 2012). To reduce admission tests' negative effect for underrepresented groups, a movement from test-mandatory to test-optional admissions has evolved in the United States (Bennett 2022; Furuta 2017), and it accelerated tremendously due to restricted testing opportunities during the COVID-19 pandemic (Camara and Mattern 2022). Regulations and implementation vary between institutions, but in general, test-optional policies do not require applicants to provide any standardized test scores (although applicants may voluntarily submit scores to be considered as part of the application). Therefore, test-optional admission policies have the potential to reduce the effect of social differences in test participation or avoidance (and not only differences in test performance).

In this study, we do not focus on SES or racial differences in test participation and test performance, however, but on gender inequalities in test-based admission as an important but underresearched topic. We investigate whether and why test-based admission policies generate a male advantage in selective college admissions: Is it due to gender differences in test performance or in test participation? Focusing on students' gender is interesting because, on the one hand, in most advanced societies, women are now more likely than men to attend college (DiPrete and Buchmann 2013) and to enroll in some prestigious fields, such as medicine (Boulis and Jacobs 2008; for Germany, see Figure A1 in the online Supplemental Material). One important reason for this is that girls perform better in school, indicated by higher grade point averages (GPAs), on average (DiPrete and Buchmann 2013). On the other hand, empirical evidence points to a male advantage in test performance, especially when tests rely on math, science, and quantitative skills (Buchmann, DiPrete, and McDaniel 2008). Hence, male students' higher scores in such tests might somewhat compensate for their lower average school performance (Azen, Bronner, and Gafni 2002; Saygin 2020). Moreover, experimental

studies suggest women avoid competitive test situations more than men do (Niederle and Vesterlund 2011); men might therefore be more likely to participate in admission tests. Thus, optional (instead of mandatory) tests could increase (rather than reduce) gender differences in test-based admissions.

Against this background, with regard to gender (and in contrast to social class and ethnic minorities), the "to-increase-diversity" argument—usually directed against test-based admissions—could be made in *favor* of test-based admissions. Whereas dis/advantages in admission tests and school performance—the two most prominent "merit"-based admission criteria—typically point in the same direction for students from a similar social class and race (Zwick 2019), they could point in opposite directions for male and female students, potentially neutralizing group-specific advantages in admission criteria. Therefore, research focused on gender is needed to determine if such a neutralizing pattern exists and why—because of men's higher test performance and/or higher rate of test-taking?

Some large-scale, real-life studies on gender differences in test performance and their effect on college admissions exist (e.g., Jurajda and München 2011), but to the best of our knowledge, no representative study examines test avoidance as a source for male advantage in test-based admissions. One reason for this lack of research is that differences in test avoidance and their effect on admission chances can only be assessed if not participating in tests is a real option for applicants. However, prior studies were mostly conducted in contexts with (quasi) test-mandatory admissions. When test scores are *required* as part of the application, studying test avoidance is not possible because not participating in tests is confounded with not applying at all; it is thus not observable from application data.

This problem also occurs with test-optional admissions if their alternative are test-mandatory admissions, as in the United States. First, the majority of (selective) colleges still require tests or recommend reporting test scores, which is mirrored by students' constantly high test-taking rates.¹ Second, admission officers consistently rate test scores among the most important admission criteria (Rosinger, Ford, and Choi 2021). Third, the test scores used for college admissions are sometimes required for other purposes (e.g., for course placement decisions; Bennett 2022;

Zwick 2019). Consequently, a large share of applicants continue to send test scores to test-optional colleges and universities (Summit Educational Group 2021).

Two studies have examined the effect of test-optional admissions on the gender composition of applicants and enrolled students in the United States. Bennett (2022) found that de-emphasizing test scores increased the share of female enrollees by 4 percentage points overall (based on about 100 private higher education institutions that introduced test-optional admission policies). In contrast, Saboe and Terrizzi (2019) did not find an effect of test-optional on the share of male applicants, based on a much larger sample of almost 1,800 four-year colleges (7 percent with test-optional policy). However, in both studies, test-mandatory admission was the alternative to test-optional admission, and thus the free choice of test participation was suppressed.

In our study, we exploit the unique German context of test-optional admission to medical schools (for more details, see the “Institutional Context” section) and use novel register data from the German central admission system for medical schools for the entire cohorts of applicants from 2012 to 2018 (see the “Data and Methods” section). Here, the alternative admission procedure to test-optional is not test-mandatory but test-free. Quite a few medical programs in Germany are test-free, that is, they do not consider test scores at all (25–50 percent between 2012 and 2018, see Figure D1 in the online Supplemental Material), and all programs that use admission tests (hereafter, test-based programs) are test-optional. Thus, German applicants have a free choice of whether to participate in admission tests because not taking these tests does not disqualify them from applying for medicine—a crucial precondition for studying test avoidance. Moreover, medicine is one of the most selective fields of study in Germany (Finger 2022) and the only field with large-scale use of a standardized admission test. These factors offer a unique opportunity to study gender differences in both test participation (or avoidance) and test performance and their effect on the chances of admission.

Our main finding is a male advantage in admission chances (among applicants with a similar GPA on their school leaving certificate), which mainly results from male applicants’ higher test performance and, to a lesser extent, from a higher incidence of female applicants’ test-avoidant

behavior. However, the effect of both depends on GPA: Male applicants’ higher test performance emerges only among test-takers with a lower GPA, and female applicants’ higher test avoidance occurs only for those with a medium GPA. Ultimately, we find that the female advantage in school performance and the male advantage in test-based admissions almost neutralize each other with respect to overall admission chances.

PREVIOUS RESEARCH AND THEORETICAL CONSIDERATIONS

We consider two research strands to inform our theoretical arguments. First, we turn to research, mainly from psychology and economics, that relies on lab experiments to analyze gender differences in test participation or performance (e.g., Cahlíková, Cingl, and Levely 2020; Lovaglia et al. 1998; Niederle and Vesterlund 2011; Spencer, Steele, and Quinn 1999). These studies are able to isolate important mechanisms behind these differences, such as stereotype threat, biased self-assessment, or competitiveness, but they lack the external validity and relevance of high-stakes admission tests.

Second, we include studies that examine the performance of men and women in high-stakes tests, with some also examining how this relates to college admissions (e.g., Azen et al. 2002; Jurajda and München 2011; Mau and Lynn 2001; Nankervis 2011; Ors, Palomino, and Peyrache 2013; Saygin 2020; Zhang and Tsang 2015). These studies focus on gender differences in test performance but do not consider (or are often unable to identify) differences in test participation.

Gender Differences in Test Participation

Standardized admission tests are highly competitive and mixed-sex situations. If tests cover science knowledge, quantitative skills, and logical reasoning, as is usually the case in admission tests for medical schools (Leiner, Scherndl, and Ortner 2018; Stumpf and Jackson 1994; Tiffin et al. 2014), gender will likely be a salient status characteristic. In such a situation, stereotype threat, (biased) self-assessment, competitiveness, and test anxiety might result in male advantages in test participation and test performance.

Concerning test participation, lab experiments with college students show that women are more likely to shy away from competitive (especially mixed-sex) test situations than are equally competent men (Booth and Nolen 2012; Gneezy, Niederle, and Rustichini 2003; Niederle and Vesterlund 2011). As potential mechanisms, Niederle and Vesterlund (2011) identify men's stronger preferences for competition and their overconfidence in their abilities. The latter means men are more likely to believe they lead the ranking than are equally well-performing women. Moreover, because of stereotype threat—that is, the “experience of being in a situation where one faces judgement based on societal stereotypes about one's group” (Spencer et al. 1999:5)—women might be more likely to “avoid the evaluative threat” by disidentifying with test situations in which math and science skills are required (Spencer et al. 1999:7). Likewise, Correll (2001, 2004) finds that cultural beliefs about gendered competencies in mathematics lead to biased self-assessment and, consequently, gendered aspirations and decisions. If a social category is salient in a given situation, performance expectations are judged more critically by individuals from the low-status group (i.e., women in math- and science-based tests) and more leniently by high-status individuals. Thus, at a given ability level, men are more likely to overestimate their actual task ability and women more likely to underestimate it (Correll 2004; Penner and Willer 2019). Stereotype threat and biased self-assessment are often discussed as reasons for higher levels of test anxiety among low-status individuals (Lovaglia et al. 1998; Spencer et al. 1999; Zeidner 1990), and they might also contribute to more test avoidance among women.

Differences in school performance are another potential source of gendered test-taking decisions. Universities often rely on a combination of GPA and test scores to select applicants, which they sometimes combine with additional criteria (Azen et al. 2002; Saygin 2020). It is well established that girls outperform boys in school (e.g., DiPrete and Buchmann 2013; Voyer and Voyer 2014). Thus, lower participation rates for women in college admission tests may just be a rational decision: Women might simply less often feel the need to invest in test-taking efforts (e.g., preparation, travel to the test site, and fees) because they can more often rely on their high GPA.

Male applicants, by contrast, might participate in tests more often (and invest in test preparation) to compensate for their, on average, lower GPA.

We cannot test empirically for the different theoretical behavioral accounts of competitiveness, stereotype threat, biased self-assessment, or test anxiety. We can, however, investigate their relevance versus the (need to compensate) argument. If gender differences in test participation are caused by gender differences in the need to compensate for insufficient school performance (i.e., compositional differences in GPA), we should not observe gender differences in test avoidance after controlling for GPA. If differences exist after controlling for GPA, gender differences in test participation are plausible, at least in part, due to the aforementioned theoretical behavioral accounts. Based on these considerations, we hypothesize the following:

Hypothesis 1: Female applicants are less likely than male applicants to participate in standardized admission tests.

Hypothesis 2: The gender difference in test participation is smaller when controlling for GPA than without this control (gender differences in the “need to compensate” argument).

The remaining gender effect after controlling for GPA can be taken as an indication of behavioral differences in women's and men's test-taking decisions.

We can also expect a nonlinear relationship between GPA and test participation in interaction with gender. Concerning nonlinearity, applicants with a medium GPA have the most to gain from good test results: Good test results can help increase their position in the ranking (compared to GPA only). In contrast, the initial ranking position of applicants with comparatively poor grades is quite low, and good test results (if they manage to get them) will only lift them to the middle of the distribution, which would still not be sufficient for admission. This suggests applicants with a medium GPA are more likely to participate in tests than are applicants with a very good GPA (who do not need tests to secure a high ranking) or a very poor GPA (who are not very likely to achieve a promising ranking through tests).

This concave (inverse U-shaped) association between GPA and test participation might be less pronounced for female students. School

grades as external assessments of previous performance are more important for the self-assessment of female students' math competence because they help counterbalance cultural beliefs about male superiority in this domain (Correll 2001). A medium GPA, however, might not be sufficient for girls to fulfill this counterbalancing function. In contrast, external assessments of competence seem to be less important for male applicants' self-assessment, which results in an overestimation of their actual ability. Moreover, the established inverse relationship between test anxiety and GPA is stronger for female students (Chapell et al. 2005). Thus, although the (need to compensate—most gains) argument applies to both male and female applicants with a medium GPA, the performance feedback and test anxiety mechanisms suggest this pattern is less prevalent among women than men. We therefore expect to find the following:

Hypothesis 3: Gender differences in test-taking are largest among students with a medium GPA (gendered “need to compensate—most gains” argument).

Gender Differences in Test Performance

Numerous studies show that male students outperform female students in college admission tests, especially when tests are based on math, science, and quantitative skills (e.g., Azen et al. 2002; Mau and Lynn 2001; Nankervis 2011; Saygin 2020; Zhang and Tsang 2015). This also applies to admission tests for medical programs, which are heavily science-oriented (Leiner et al. 2018; Stumpf and Jackson 1994; Tiffin et al. 2014).

Research has found gender differences in admission test scores for prestigious higher education institutions and fields of study, which constitute a high-stakes situation with elevated levels of competition and stress. In the Czech Republic, Jurajda and Münich (2011) found that female applicants (including the highest achievers among them) performed worse than men on tests for very selective universities but not on tests in less competitive situations (i.e., for less selective institutions). Similarly, Ors et al. (2013), focusing on a selective group of applicants to an elite economics program in France, found that men outperformed women in this competitive test situation

even though these men achieved lower scores in the less competitive baccalaureate exam. These findings are mirrored by experimental studies showing that men outperform women in mixed-sex tournament (vs. piece-rate) situations (Gneezy et al. 2003), when the level of stress or time pressure is high (Cahlíková et al. 2020; Shurchkov 2012), and when stereotypes or status differences are activated (Lovaglia et al. 1998; Spencer et al. 1999). An observational study on applicants to Austrian medical schools shows that test anxiety partly mediates gender differences in test performance (Leiner et al. 2018). We therefore hypothesize a male advantage in test performance:

Hypothesis 4: Male test-takers achieve higher test scores than female test-takers (*ceteris paribus* of GPA).

Again, the strength of the relationship between GPA and test performance might differ by applicants' gender for reasons similar to those discussed for test participation, such as a stronger effect of performance feedback on women's self-assessment (Correll 2001) and test anxiety (Chapell et al. 2005). This might result in a stronger negative relationship between GPA and test performance for female students. Thus, we expect the following:

Hypothesis 5: The male advantage in test performance increases with decreasing (poorer) GPAs.

In summary, the hypothesized male advantages in test participation (Hypotheses 1–3) or test performance (Hypotheses 4 and 5) should ultimately increase men's admission chances.

On a final note, one might argue that gender differences in test performance are, at least partly, generated by gender differences in the correlation of school grades and competencies. This correlation might be weaker for boys, for example, because of gender biases in teachers' grading (Protivínský and Münich 2018). However, in their meta-analysis, Voyer and Voyer (2014) found that the female advantage in school grades for the same level of competence was small to medium for language courses and very small for math courses. Thus, gender differences in test performance are less likely to indicate competencies of male applicants that are not adequately recognized in their GPA; rather, these gender

differences are more likely due to the aforementioned behavioral accounts.

INSTITUTIONAL CONTEXT

Admission System for German Medical Schools

Medical programs are the most selective programs in Germany (Finger 2022), with an average admission rate of around 25 percent. Students who want to study medicine need to apply via a central clearinghouse (Stiftung für Hochschulzulassung). Until 2019, places were allocated via three quotas: (1) a 20 percent GPA quota, (2) a 20 percent waiting-period quota, and (3) a 60 percent program-specific quota. For the last quota, GPA is a mandatory selection criterion with the highest weight, and universities can add further criteria (e.g., test scores or work experience). Hereafter, we refer to the three quotas as GPA quota, waiting quota, and university-admission quota.

Applicants can apply via one, two, or all three quotas. In each quota, they may rank up to 6 of the 35 public universities offering medical programs.² Each applicant can be admitted to only one program. The allocation mechanism is stepwise, starting with the top-ranked university in the GPA quota and closing with the lowest ranked in the university-admission quota. Applicants usually need the top GPA (1.0 on a scale from 1.0 = highest to 4.0 = lowest) or must undergo a long waiting period (approximately seven years), respectively, to be considered in the first two quotas.

Additional selection criteria in the university-admission quota potentially provide applicants with the opportunity to at least partly compensate for comparatively poorer school performance. Over the past years, admission tests have become the most frequently used additional selection criteria. The TMS (Test für Medizinische Studiengänge, or test for medical programs)—developed and run by an external organization—is the most prominent test format, used by 14 of 35 programs in 2012 and 23 programs in 2018 (see Figure D1 in the online Supplemental Material). The TMS comprises nine dimensions, including medical/scientific comprehension, quantitative/formal problems, mental rotation, learning figures/facts, and accurate/concentrated work (Kadmon and Kadmon 2016). Until 2021, the TMS could only be taken once. Test fees in 2018 amounted to

€83. Test-takers receive their results and can include them in their applications if they apply via the university-admission quota for programs that use TMS scores as a selection criterion. In our observation period, only three or four programs used further tests (hereafter, “local tests”; for more information, see Section E of the online Supplemental Material). The following analyses focus on the TMS and applications via the university-admission quota because gender differences in test participation and performance—and ultimately their effect on admission chances—can only be assessed for this quota (see the “Data and Methods” section). Figure 1 summarizes the main features of the central admission system in Germany.

Test-Optionality

Admissions to German medical schools are either test-optional or test-free. During our observation period, applicants could always choose to apply to test-free programs (18 in 2012 and 8 in 2018, see Figure D1 in the online Supplemental Material). Thus, applicants had the option to avoid taking the TMS without having to refrain from applying to medical programs or specific institutions. For admission to test-based programs, test scores are combined into a final admission score using the applicants’ GPA and further criteria, which then determines an applicant’s position in the applicant queue. The admission process is highly standardized and not biased by individual admission officers who might (un)consciously consider nonsubmission of test scores a negative signal. Thus, importantly, test scores can only increase the final admission score; submitting poor test scores or not taking the test do not result in detrimental treatment (i.e., there are no points subtracted for missing or poor test results).

Furthermore, applicants are not crowded out due to higher competition in test-free programs. Indeed, competition in test-free programs is lower than in test-optional programs: 27 versus 37 applicants per seat, on average, between 2012 and 2018 (authors’ calculations based on application register data). Thus, even if applicants misinterpret test-optional as mandatory and avoid applying to test-based programs, they have the choice of applying to less competitive test-free programs. Additionally, German universities (including medical programs) are rather homogeneous in terms of

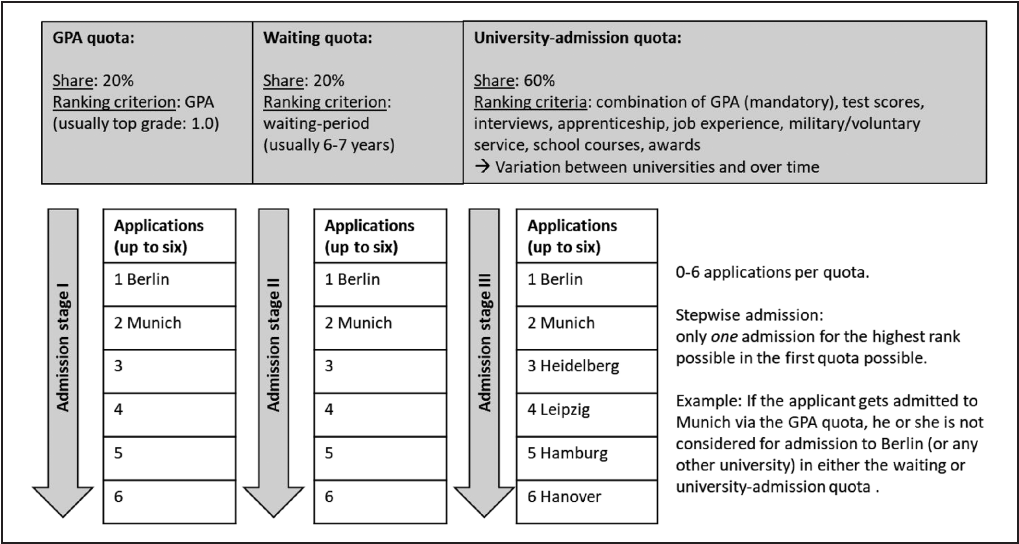


Figure 1. Central application and admission system for medical programs in Germany.
Source: Authors' own depiction.
Note: Admission system until 2019. Locations of medical programs listed in the figure are examples.

quality and prestige (Mayer, Müller, and Pollak 2007). Thus, the use of test-based admission policies is only weakly correlated with institutional prestige.³ These characteristics of the German admission system to medical schools provide ideal preconditions for studying gender differences not only in test performance (as can be done in other test-based contexts) but also in test participation and the effect of both on admission chances.

DATA AND METHODS

Data

We use individual-level application register data provided by the central clearinghouse that cover the whole population of applicants to medical programs at German public universities for the winter terms 2012 to 2018, amounting to around 300,000 applicants.⁴ Comparable data for other study programs, which are not allocated centrally, do not exist in Germany.

The data contain (a) applicants' application pattern (ranking of universities in each quota); (b) information on whether, where, and via which quota applicants were admitted; (c) applicants' GPA as a universal admission criterion; (d) some sociodemographic characteristics, such as gender

and age (the data do not contain any information on applicants' SES or ethnicity because they are not used in the admission process); (e) information on applicants' school history (year, German state, and type of higher education entrance certificate); and (f) performance-related information if used as selection criteria by the ranked programs in the university-admission quota. Thus, TMS scores are only included when applicants ranked at least one TMS-based program in this quota and reported their scores. With regard to local tests, we only know whether applicants applied to a local-test-based program but neither their actual test participation nor test scores. We therefore do not include them in the main analyses but provide sensitivity analyses with hypothetical test-participation assignment (see Table 1 and the "Results" section). We added yearly information on whether the 35 programs used admission tests as a criterion in the university-admission quota to the individual-level data (for the list of sources, see Section B of the online Supplemental Material).

Variables and Analytic Strategy

Our analytic strategy includes three steps. We examine gender differences in (1) test participation and (2) test performance, and we (3) analyze

Table 1. Description of Sample Restrictions and Study Samples.

Exclusion rule	Total			Test participation		Test performance		Admission	
	All applicants to medical programs, 2012–2018 ^a	Excluding those without applications in university-admission quota		Additionally, excluding those who applied to a local-test-based program and did not report TMS scores = study sample		Additionally, excluding those who did not report TMS scores = study sample		Study sample “test participation” plus excluding those with admission approval in the GPA or waiting quota = study sample for effect of test participation	Additionally, excluding those who did not report TMS scores = study sample for effect of test performance
N	293,299	265,176		223,621		71,187		207,872	67,264
Share of men (%)	35.4	34.9		34.8		35.2		34.6	35.1
Average GPA (SD) ^b	1.97 (0.62)	1.95 (0.61)		1.95 (0.61)		1.84 (0.50)		1.95 (0.58)	1.84 (0.48)
Male average GPA (SD)	2.07 (0.65)	2.03 (0.64)		2.03 (0.63)		1.91 (0.52)		2.04 (0.61)	1.91 (0.50)
Female average GPA (SD)	1.92 (0.60)	1.90 (0.59)		1.90 (0.57)		1.80 (0.48)		1.91 (0.56)	1.80 (0.46)
Sensitivity analyses ^c	Including applicants who only applied via GPA and/or waiting quota: hypothetical assignment with test participation = 0 (lower bound) and = 1 (upper bound)								
	Including applicants who applied to local-test-based programs and did not report TMS scores: hypothetical assignment with test participation = 0 (lower bound) and = 1 (upper bound)								
	Estimates based on the “admission samples” (consistent sample restrictions for all estimates)								

Source: Stiftung für Hochschulzulassung application register 2012 to 2018, authors’ calculation.

Note: TMS = Test für Medizinische Studiengänge (test for medical programs); GPA = grade point average.

^aNot included are applicants who had missing information (0.05 percent), who were admitted via a special military quota (0.5 percent), and who already graduated from a higher education program (2.9 percent).

^bGerman GPA ranges from 1.0 (highest) to 4.0 (lowest).

^cFurther information on the sensitivity analyses are presented in the online Supplemental Material, Section C (rational), Section E (test participation), and Section F (test performance).

their consequences for gender differences in admission chances. Although we use the register data covering the entire population, we report conventional significance levels (a) because we had to exclude some cases with missing information (see next subsection) and (b) to ascertain that the gender differences found are not due to chance (Rubin 1985).

First, to test our hypotheses on test participation (Hypotheses 1–3), we estimated logistic regression models and reported average marginal effects (AMEs) to facilitate comparison of effect sizes across models (Breen, Karlson, and Holm 2018). Gender and GPA categories, including their interaction, served as our main independent variables. All models control for age, German state of high school graduation, and year of application.⁵ Information on whether or not applicants *reported TMS results* served as the dependent variable. Because test-taking and reporting of test scores are optional, it is not sufficient to take applications to test-based programs as a measure for test participation. Accordingly, the share of applicants reporting TMS scores (about 30 percent) is much lower than the share of applicants applying to at least one TMS-based program (about 90 percent) in the university-admission quota.

Whether applicants participated in the TMS but did not report their test scores is not observable with the register data. It could nevertheless bias our results if nonreporting of test scores differs between male and female applicants. Even though this is not a rational behavior, because test scores can only improve applicants' admission scores, we cannot entirely rule out underreporting and gender differences. This would be in line with the double standard of self-assessment (Correll 2004) introduced in the theory section. To assess the extent of potential bias, we provide additional analyses based on an alternative, but otherwise limited, data source (for details, see the "Results" section and Section H of the online Supplemental Material).

Second, to test our hypotheses on test performance (Hypotheses 4 and 5), we focus on test-takers (i.e., applicants who reported TMS test scores). Their *test scores* are the dependent variable. TMS test scores range (continuously) from 1.0 (highest) to 4.0 (lowest). We inverted the scale so that higher values indicate better test performance. We estimated ordinary least squares (OLS) regressions using the same independent and control variables as for test-taking (see previous discussion).

Again, gender differences in nonreporting might bias our results if gender differences in the probability of reporting depend on the level of test scores. For that reason, we again provide additional analyses to identify and quantify a potential bias (see the "Results" section).

Finally, to assess the importance of gender differences in test participation and test performance, we examine their role in successful applications. We estimated logistic regression models (reporting AMEs) using *admission in the university-admission quota* as the dependent variable (because admission tests are only relevant in this quota) and our measures of test participation and performance as the main independent variables. The estimations include gender, GPA categories, the control variables mentioned previously, and the number of applications per applicant because more applications could increase admission chances.

Sample Definition

Due to the different quotas and the stepwise admission procedure (see the "Institutional Context" section), it would be neither possible nor meaningful to run our models on the whole population. In the following, we define the study samples for our three analytic steps (for further information on our sample restrictions, see Section C of the online Supplemental Material). As a preparatory step, we listwise deleted some cases with missing information on variables (i.e., not reported by the applicants; $n = 145$), starting with 293,299 applicants.

The study sample for *test participation* consists of 223,621 applicants. We first restricted the sample to applicants who applied via the university-admission quota because tests—our central variable—are only used as part of this quota (thus, information on test participation and performance is only available for these applicants). This restriction is potentially problematic if only applying via the GPA and waiting quota indicates test-avoidant behavior. We therefore conducted a sensitivity analysis by including applicants who only applied via the GPA and waiting quotas and categorizing them as test-avoiders (see the "Results" section).

Second, we excluded applicants who applied to a local-test-based program and did not report TMS scores because they cannot easily be categorized as either test-takers or test-avoiders (see

Table 2. Gender and GPA Effects on Test Participation (TMS Scores Reported).

	Model 1	Model 2
Gender: male	0.003 (0.002)	0.020*** (0.002)
GPA categories, reference = 1.0–1.1 (highest)		
1.2–1.3		0.024*** (0.005)
1.4–1.5		0.109*** (0.005)
1.6–1.7		0.137*** (0.004)
1.8–1.9		0.102*** (0.004)
2.0–2.4		–0.007 (0.004)
2.5–2.9		–0.125*** (0.004)
3.0–4.0 (lowest)		–0.218*** (0.004)
Control variables	Yes	Yes
Akaike information criterion	273,568	262,113
Bayesian information criterion	273,825	262,443
R ² (Maddala)	0.030	0.079

Sources: Stiftung für Hochschulzulassung application register 2012 to 2018, study sample “test participation” (see Table 1), Stiftung für Hochschulzulassung annual publications (see Section B of the online Supplemental Material), authors’ calculation.

Note: Average marginal effects based on logistic regressions; standard errors are in parentheses. $n = 223,621$. For the control variables included, see the “Data and Methods” section. For the full models displaying all control variables, see Section E of the online Supplemental Material. TMS = Test für Medizinische Studiengänge (test for medical programs); GPA = grade point average.

*** $p < .001$ (two-tailed test).

explanation in the data subsection and Section E of the online Supplemental Material; $n = 41,555$, 14.2 percent of applicants). As a sensitivity analysis, we reestimated the models of test participation by categorizing applicants to local-test-based programs (who did not report TMS scores) as test-participants and test-avoiders to receive upper- and lower-bound estimates.

The study sample for *test performance* consists of 71,187 test-takers, defined as applicants who participated in the TMS and reported their scores to the clearinghouse. They are the only cases with information on test performance.

Finally, in our analyses on *admission chances*, we additionally excluded applicants who were admitted via the GPA or waiting quota. Because each applicant can gain admission to only one program (with a stepwise progression starting with the GPA quota; see Figure 1), for these applicants, test participation and test performance do not have any effect on their admission. This further restriction leaves us with 207,872 and 67,264 applicants, respectively, for estimating the effects of test participation and test performance on admission chances. As a sensitivity analysis, we reestimated all analyses on test participation and test performance with these most restricted samples.

Table 1 summarizes the sample restrictions, the resulting compositions of the different study samples, and the different sensitivity analyses. Further descriptive information on central variables used in the different study samples (differentiated by gender) are presented in Section D of the online Supplemental Material.

RESULTS

Test Participation

We begin by testing Hypotheses 1 to 3 on test participation. On average, 32 percent of the applicants in the university-admission quota reported test scores (see Table D1 in the online Supplemental Material). Table 2 displays the AMEs for gender and GPA on “test scores reported” (our measure of test participation), based on logistic regressions. Model 1 shows that test participation does not differ between male and female applicants, contradicting Hypothesis 1.

By adding GPA in Model 2, the gender coefficient increases to only 2 percentage points. Because the overall test-participation rate is 32 percent, this means a 6 percent increase in relative terms ($2 / 32 = 0.06 \times 100$). Thus, male

applicants are slightly more likely to participate in tests than are female applicants with the same GPA. This finding also contradicts Hypothesis 2. To support this hypothesis, we would have needed to see a positive effect of being male in Model 1 to be “compensated” (i.e., reduced) in Model 2. The finding that gender differences in test participation only occur when controlling for GPA indicates, however, that behavioral explanations, rather than GPA-compositional explanations (i.e., gender differences in the “need to compensate for poorer GPAs”), are at work.

We further differentiated Model 2 by the number of applications to test-based programs (zero to two, three to four, and five to six; with 0 = applicants only ranked non-test-based programs in the university-admission quota and 6 = applicants exploit the maximum number of test-based programs). The rationale for this analysis is that the likelihood of test participation might be positively associated with the number of applications to test-based programs (to enhance admission chances), and this strategic use of test scores might be more often executed by men—possibly motivated by their greater confidence in their own performance (Correll 2001; Niederle and Vesterlund 2011; Penner and Willer 2019). Analyses, presented in the online Supplement Material (Table E1), show that the slightly higher probability of male test participation, displayed in Model 2, is indeed driven by individuals who applied for five to six test-based programs: here, men are 4.3 percentage points more likely to participate in tests than women (three to four test-based programs = 1.2 percentage points; zero to two test-based programs = 0.1 percentage points).

Hypothesis 3 postulates a nonlinear effect of GPA on gender differences in test participation. Model 2 (in Table 2) shows the expected concave association between GPA and test participation. Compared to the reference group with a top GPA, the probability of participation increases until a medium GPA (13.7 percentage points higher for those with a GPA of 1.6–1.7) and then decreases sharply again. This indicates that test scores are used to compensate for average school performance—and are less frequently used by individuals with a very good GPA (for whom tests are not necessary to boost their admission chances) and those with a poor GPA (for whom test scores are unlikely to achieve

a promising admission score). The latter more often apply via the waiting quota to improve their admission chances.⁶

To test Hypothesis 3, we included interaction terms between gender and GPA categories in the regressions, visualized as predicted probabilities in Figure 2. This figure shows, first, quite similar patterns for male and female applicants. Second, medium-performing (male and female) applicants to five and six test-based programs are most likely to report TMS scores. Third, in this group, gender differences are statistically significant in the GPA categories between 1.4 and 2.9 (see Figure E1 in the online Supplemental Material), with a size of up to 5.7 percentage points in the 1.6 to 1.9 GPA categories. Interestingly, the share of male applicants does not differ by the number of applications to TMS-based programs, but only by whether they report test scores. These findings support Hypothesis 3.

In summary, our results on gender differences in test participation support what would be expected if the proposed mechanisms of gender differences in test anxiety, confidence, and stereotype threat were true. It is important to note, however, that in relative terms, applicants' GPAs are more predictive than their gender. Finally, we cannot rule out that applicants who took the test but achieved poor test results are less likely to report their scores (see sensitivity analyses in the following for an exploration of potential biases).

Test Performance

Figure 3 presents the distribution of GPA and test scores by gender, indicating a female advantage in GPA and a male advantage in test scores. This lends initial support to Hypothesis 4, that male test-takers achieve higher test scores than women.

Table 3 displays OLS regressions on the relationship between gender, GPA, and test scores. Positive coefficients indicate better test performance. Model 1 (and Model 2 controlled for GPA) shows that male test-takers score higher than their female peers, which supports Hypothesis 4. The male advantage of 0.163 test points (and 0.183, respectively) seems to be small but is in fact substantial, signifying an increase in admission chances by about 4.8 (5.4) percentage points (or 15 [17] percent relative to the admission rate of test-takers of 31.4 percent).⁷

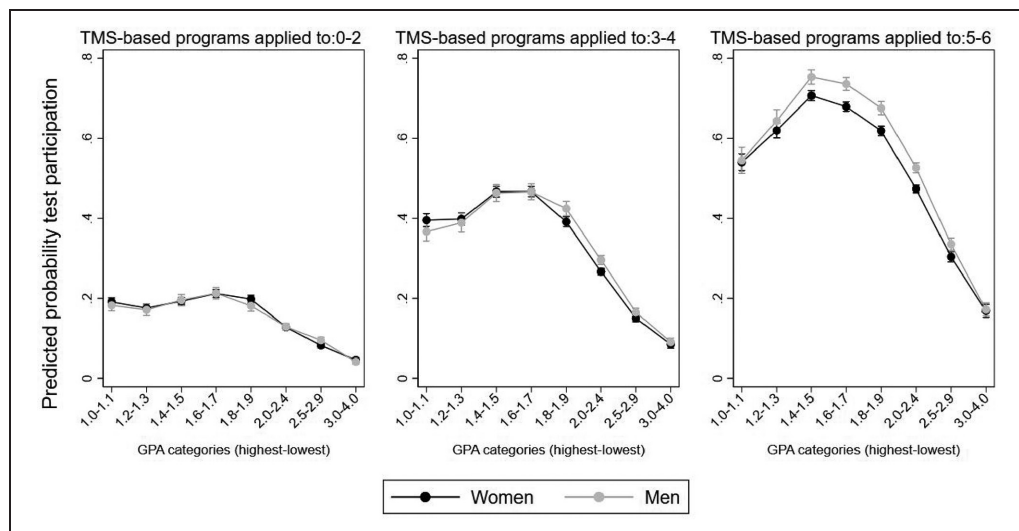


Figure 2. Predicted probabilities of test participation by gender and GPA categories (95% confidence intervals). Sources: Stiftung für Hochschulzulassung application register 2012 to 2018, study sample “test participation” (see Table 1), Stiftung für Hochschulzulassung annual publications (see Section B of the online Supplemental Material), authors’ calculations.

Note: Predicted probabilities based on logistic regression models with interaction terms gender \times GPA categories (see Section E of the online Supplemental Material). GPA = grade point average.

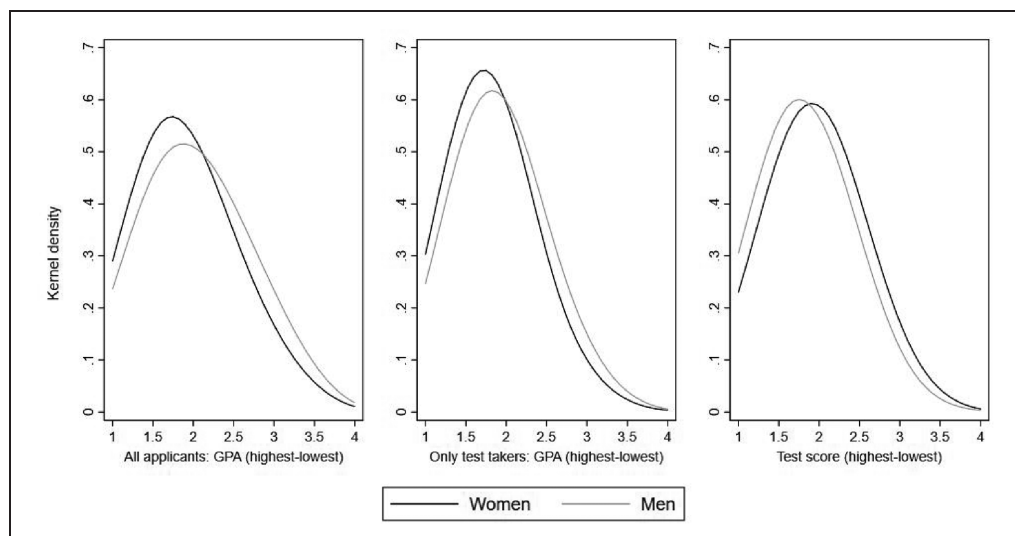


Figure 3. Distribution of grade point average (GPA) and test scores by gender.

Source: Stiftung für Hochschulzulassung application register 2012 to 2018, study samples “test participation” (left) and “test performance” (middle and right; see Table 1), authors’ calculations.

Table 3. Gender and GPA Effects on Test Performance (Test-Takers Only).

	Model 1	Model 2	Model 3
Gender: male	0.163*** (0.004)	0.183*** (0.004)	0.116*** (0.015)
GPA categories, reference = 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	Yes	Yes	Yes
Constant	3.933*** (0.017)	3.499*** (0.018)	3.524*** (0.019)
R ²	0.106	0.192	0.194

Source: Stiftung für Hochschulzulassung application register 2012 to 2018, study sample “test performance” (see Table 1), authors’ calculation.

Note: Ordinary least square models; standard errors are in parentheses. *n* = 71,187 applicants reporting test scores. Inverted test scores (1 = low, 4 = high). For the control variables included, see the “Data and Methods” section. For full models, see Section F of the online Supplemental Material. GPA = grade point average.

****p* < .001 (two-tailed test).

The estimates in Model 3, including interaction terms of gender and GPA categories, confirm Hypothesis 5: The male advantage in test performance increases with lower GPA. Gender differences do not emerge for test-takers with an excellent or very good GPA, whereas male test-takers with a GPA between 1.6 and 4.0 score higher than their female counterparts. A potential explanation is women’s stronger reliance on external feedback (Correll 2001): Women with a high GPA have already demonstrated strong math and science skills and received (grading) feedback. As a result, they might be less affected by test anxiety (see the “Theory” section).

Admissions

Finally, we assess the consequences of test participation and test performance for gender differences in admission chances. Models 1 to 3 in Table 4 focus on the role of test participation, and Models 4 to 6 show the role of test performance. The latter

models are restricted to applicants who reported test scores. Model 1, which only includes gender and control variables, displays no gender difference in admission chances. After including GPA in Model 2, male applicants are 2.8 percentage points more likely to be admitted than female applicants, indicating a male advantage despite the lower average GPA of male applicants (see also Figure 3). This effect might seem small in absolute terms. However, because the overall chance of admission only amounts to 22 percent (see Table D3 in the online Supplemental Material), this is a substantial difference of about 13 percent in relative terms (2.8/22 = 0.127 × 100). As expected, Model 2 shows a strong negative association between GPA and admission chances, with admission chances of almost zero for those in the bottom GPA category.

According to Model 3, TMS participation increases the admission probability by 12.2 percentage points but rarely decreases the male advantage in admission chances. To examine whether male applicants benefit more strongly

Table 4. Effects of Gender, Test Participation, and Test Performance on Admission.

	Test participation (TMS scores reported)			Test performance (test-takers only)		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
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, reference = 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-standardized)			0.155*** (0.001)			
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Akaike information criterion	203.847	99.530	90.236	74.142	46.337	32.583
Bayesian information criterion	204.113	99.848	90.564	74.379	46.620	32.874
R ² (Maddala)	0.059	0.431	0.455	0.133	0.427	0.533

Sources: Stiftung für Hochschulzulassung application register 2012 to 2018, study samples for admission (see Table 1), Stiftung für Hochschulzulassung annual publications (see Section B of the online Supplemental Material), authors' calculation.

Note: Average marginal effects based on logistic regressions; standard errors are in parentheses. n (test participation) = 207,872; n (test performance) = 67,264. For the control variables included, see the "Data and Methods" section. For full models, see the online Supplemental Material, Section G, Table F1. TMS = Test für Medizinische Studiengänge (test for medical programs).

*** $p < .001$ (two-tailed test).

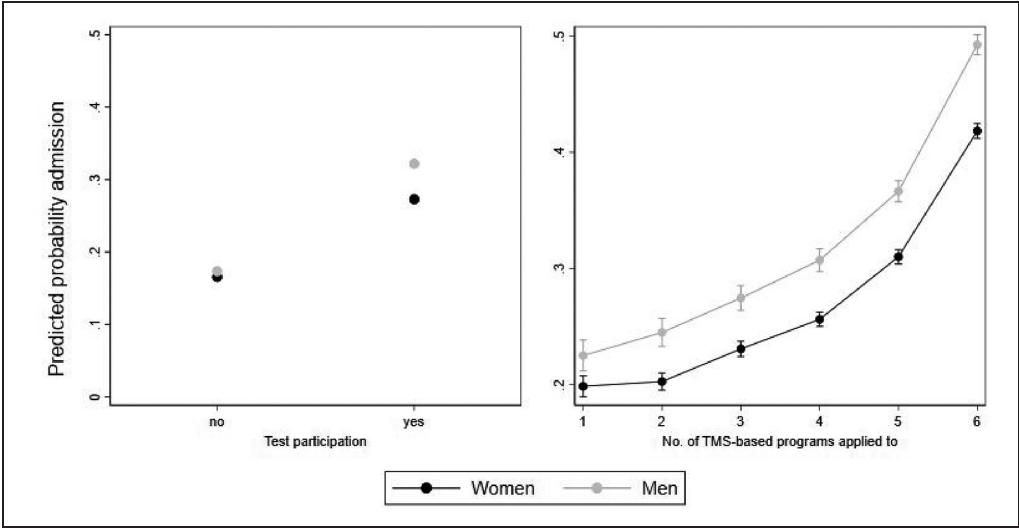


Figure 4. Predicted probabilities of admission by gender, test participation, and number of TMS-based programs applied to (95% confidence intervals).
Sources: Stiftung für Hochschulzulassung application register 2012 to 2018, study sample “test participation” for admission (see Table 1), Stiftung für Hochschulzulassung annual publications (see Section B of the online Supplemental Material), authors’ calculation.
Note: Predicted probabilities based on logistic regression models with interaction terms Gender \times Test Participation/Number of Applications to TMS-based programs. Right panel only includes test-takers. For the control variables included, see the “Data and Methods” section. For full models, see Table G2 in the online Supplemental Material. TMS = Test für Medizinische Studiengänge (test for medical programs).

from test participation, we added interaction terms between test participation and gender to Model 3. Figure 4 displays the results as predicted probabilities of gaining admission. The left panel shows no gender differences in admission chances among applicants who did not report their test scores. In contrast, male applicants reporting their TMS scores were 5 percentage points more likely to gain admission than female applicants.

The right panel of Figure 4 focuses on test-takers and explores whether taking tests increases admission chances, especially if many TMS-based programs are ranked. The rationale is to assess whether the (positive) effect of test participation differs between application strategies. Among test-takers, applying to a high number of TMS-based programs indeed increases admission chances substantially. The gender gap in admission chances grows slightly with each additional application to TMS-based programs and amounts to 7.5 percentage points among individuals who exploit the maximum number of applications to test-based programs. The question, however, is whether these higher admission chances are solely

due to men’s somewhat higher rates of test participation (reported in Table 2) or due to better test performance, which we analyze next.

Model 4 and Model 5 (in Table 4) replicate Model 1 and Model 2 for the sample of test-takers. Gender differences in the admission probability are larger among test-takers. Increases in the gender coefficient between Model 4 and Model 5 are similar to increases between Model 1 and Model 2. Thus, within the sample of test-takers, there seems to be a male advantage in admission chances. With Model 6, we assess the effect of the male advantage in test performance (previously described) on admission chances. We find that an increase in the TMS score by 1 SD (which is 0.53, see Table D2 in the online Supplemental Material) increases the likelihood of admission by 15.5 percentage points. This is a substantial increase of almost 50 percent relative to test-takers’ average admission rate of 31.4 percent (see Table D3 in the online Supplemental Material). Moreover, comparing the gender coefficients between Model 5 and Model 6, we observe a sharp drop—from 6.3 to only 1.6 percentage points—in

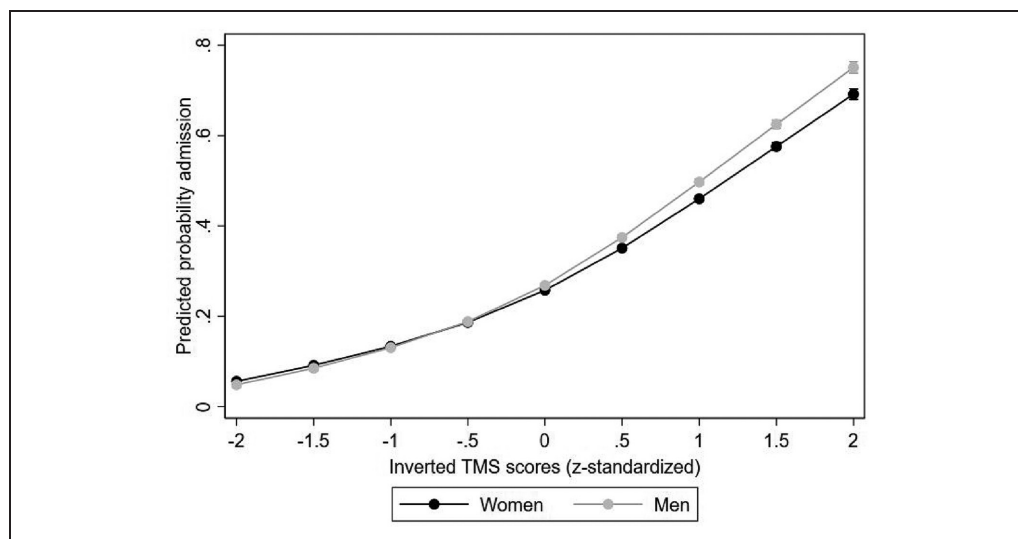


Figure 5. Predicted probabilities of admission by gender and TMS scores (95% confidence intervals). Sources: Stiftung für Hochschulzulassung application register 2012 to 2018, study sample for admission “test-takers” (see Table 1), authors’ calculation.

Note: Predicted probabilities based on logistic regression models with interaction terms Gender \times TMS Scores. For control variables included, see the “Data and Methods” section. For the full model, see Section G of the online Supplemental Material. TMS = Test für Medizinische Studiengänge (test for medical programs).

the male advantage when including test performance in Model 6.⁸ This indicates the strong mediating role of test performance for gender differences in admission chances. Finally, looking at the remaining gender effect in Model 3 (0.025) and Model 6 (0.016), the female advantage in GPA and the male advantage in test scores seem to almost neutralize each other, leading to nearly equal chances of admission via the university-admission quota. Note that this conclusion on ultimate gender parity in admission chances is not counteracted by admissions via the other two quotas, in which gender differences are small and neutralize each other (admission via GPA quota: 3.0 [2.6] percent of female [male] applicants; via the waiting quota: 4.2 [5.8] percent of female [male] applicants).

The small remaining male advantage in admission chances (Model 3 and Model 6) might be due to unobserved characteristics—such as medical-related work experience (e.g., because of alternative military service performed in a care facility), which might be more prevalent among high-performing men—rather than discrimination. Figure 5 indicates the male advantage slightly increases with test performance, and it is not clear why only

high-performing women (and not low-performing women) should be discriminated against compared to their male competitors.

Sensitivity and Additional Analyses

We performed several sensitivity analyses based on different sample definitions and measurements of our dependent variable for test participation (see the “Data and Methods” section). Respective tables and figures are presented in the online Supplemental Material.

First, to provide all estimates based on one consistent sample, we recalculated all models for test participation and performance based on the same sample as our analyses of admission chances, that is, without individuals who were admitted via the GPA or waiting quota. The findings are very similar to those reported here (see Table E6 and Figure E4 for test participation, Table F2 for test performance, in the online Supplemental Material).

Second, we (re)included certain applicant groups in the sample and categorized (a) those who only applied via the GPA and/or waiting

quota as test-avoiders and those who applied to a local-test-based program (but did not report TMS scores) as either (b) test-avoiders (lower-bound estimates) or (c) test-takers (upper-bound estimates; see Tables E3–E5 and Figures E2 and E3 in the online Supplemental Material). The findings are substantially similar to the results of our main analyses.

Third, we approximated test participation by “test-score reporting.” Even though it is not a rational behavior to withhold test scores in the German context, we cannot rule out underreporting. Our conclusion on gender differences in test participation and performance could be incorrect if male and female students were differently likely to report (specific) test scores. With the register data, we are only able to observe test scores for participants who chose to submit them. We also conducted an online survey of the 2018 applicant cohort that contains information on test participation. We use this data set and compare the information on test participation (weighted based on the register data 2018) with our measure of “test-score reporting” in the 2018 register data (see Table 5). Details on the online survey data are provided in Section H of the online Supplemental Material.⁹

In 2018, 97 percent of applicants applied to at least one test-based program, but as Table 5 (upper part) shows, only around 54 percent took the test (online survey). Moreover, only around 34 percent of 2018 applicants reported test scores (register data). This illustrates that test-taking is truly optional and is interpreted as such in the German context. The overall 20-percentage-point gap between participation and reporting is equally high for men and women.

The distribution of test-score reporters across the GPA categories is very similar for male and female applicants. In contrast, gender differences do occur with regard to test participation by GPA. In general, applicants with medium school performance—that is, those with the highest compensatory potential—are most likely to participate in tests. However, among test-takers with top GPAs (and thus among applicants who often do not need to compensate for their GPA), women are slightly more likely than men to participate in tests (46.9 percent vs. 40.8 percent) but also more likely to underreport their test scores (16.3 percentage points vs. 9.5 percentage points). In contrast, men are more likely to participate when

their GPAs are midrange (72.6 percent vs. 68.0 percent) and poor (44.4 percent vs. 41.7 percent). This is again in line with the theoretical account that women rely more strongly on external feedback (Correll 2001).

The lower part of Table 5 shows the male advantage in test performance (mean TMS scores) is very similar among test-reporters and test-takers. The same applies to gender differences within the GPA categories. In summary, our main analyses on test participation (taking test-score reported as the indicator) are unlikely to be disturbed by gender differences in nonreporting.

DISCUSSION

Before drawing conclusions from our findings, we contextualize our study with regard to sex segregation in fields of study and intersectionality (i.e., gender and SES). Medical programs are special in that they are female-dominated *and* have a strongly science-based curriculum (Barone 2011), which is reflected in admission tests to medical schools. In contrast, admission tests for typically female-dominated programs (e.g., the humanities) more often rely on qualitative and verbal contents—a female-typed domain—so that negative stereotypes of female inferiority should not be activated. Thus, here a male advantage in test-based admissions is likely less pronounced or even reversed (see Correll 2001).

The opposite could be expected for male-dominated fields, such as engineering. Here, the male advantage in test-based admissions might be even more pronounced than in our study. Not only are the content of the test and the field stereotypically male, but the testing situation is also more likely to be dominated by male test-takers. If so, admission tests would further decrease the gender diversity of the student body, in contrast to female-dominated medical programs. On the other hand, female students applying for male-dominated fields might be much more positively selected regarding their math/science grades and related competencies, on average, compared to female applicants in other fields and male competitors. Thus, gender differences in test participation and performance might be less pronounced here than in medical programs. Future research should thus compare gender inequality in test-based admissions between gender-balanced, male-dominated, and female-dominated fields.

Table 5. Comparison of Test-Score Reporting and Test Participation.

	Register data 2018		Online survey of 2018 applicants	
	% who reported TMS scores		% who participated in the TMS	
	Men	Women	Men	Women
Total	33.6	34.2	53.5	53.6
% within GPA categories				
1.0–1.3 (top)	31.3	30.6	40.8	46.9
1.4–1.9	46.8	45.3	72.6	68.0
2.0–4.0	25.1	24.5	44.4	41.7
N (study sample “test participation”)	9,948	21,181	971	2,716
Mean (SD) of TMS scores (inverted) of				
	Test reporters		Test participants	
	Men	Women	Men	Women
Total	3.20 (0.52)	3.10 (0.53)	3.02 (0.58)	2.95 (0.58)
Means within GPA categories				
1.0–1.3 (top)	3.43 (0.47)	3.33 (0.49)	3.19 (0.56)	3.14 (0.55)
1.4–1.9	3.24 (0.50)	3.14 (0.50)	3.03 (0.56)	2.98 (0.58)
2.0–4.0	3.04 (0.51)	2.86 (0.51)	2.96 (0.57)	2.75 (0.54)
N (study sample “test performance”)	3,344	7,253	536	1,484

Source: Stiftung für Hochschulzulassung application register 2018, online survey of applicants 2018 (Waves 1 and 2), authors' calculations.

Note: GPA categories are further aggregated to ensure sufficient case numbers, especially for male applicants in the online survey data. Similar sample restrictions for register data and the online survey (for study sample definitions, see Table 1). Online survey numbers contain sampling weights (for details, see Section H of the online Supplemental Material). Note that the descriptive statistics on test-score reporting in Table 5 are based only on the application registers of 2018; this explains differences to our main analyses, which cover 2012 to 2018. TMS = Test für Medizinische Studiengänge (test for medical programs); GPA = grade point average.

A second issue is intersectionality. Given the female majority in medical programs, higher test participation and better test performance of the much smaller group of male applicants (only 35 percent, see Table 1) could also be caused by a higher socioeconomic selectivity of male applicants. Social background characteristics are not part of the register data. We can, however, indirectly examine this assumption. First, because school achievements (like GPAs) are known to include a substantial proportion of social background effects, our finding of a male advantage only among applicants with similar GPAs contradicts this assumption. Second, our data include applicants' residential postal codes (overall, Germany is divided into more than 8,000 postal codes). This allows us to add contextual sociodemographic information to the register data as approximations of the social environment in which

applicants live (and supposedly grew up). We reran our main models on test participation, performance, and admission chances including these contextual socioeconomic variables: the shares of long-term unemployed, residents with migration backgrounds, and three annual household income groups ($>€60,000$; $€30,000$ – $€60,000$; $<€30,000$). We report the results in Section I of the online Supplemental Material.

The gender coefficients do not change when including these variables, and they do not, or only very slightly, interact with applicants' gender. This suggests our findings are not biased by differences in the social selectivity of male and female applicants. Interestingly, we also find that the effects of the socioeconomic context variables are smaller than the gender effects on test participation and performance. For example, a 1 SD increase in the share of households with a high

income increases the probability of test participation by less than 1 percentage point (compared to the 2-percentage-point increase of being male).

This small effect size for social background (or rather for social environment) is not surprising when one keeps in mind that we are looking at applicants to highly prestigious study programs and control for their GPA. Thus, a high share of secondary effects of social origin (i.e., differences in educational decisions) are much less relevant at this point. Medical school applicants from lower-class backgrounds have already obtained a university entrance qualification (the *Abitur*) in the highly stratified German school system, decided to pursue a higher education degree instead of an attractive apprenticeship (Powell and Solga 2011), and chosen to apply for this highly competitive program.

CONCLUSIONS

Admission test policies in higher education are highly debated and undergoing rapid changes. One major criticism is that admission tests further disadvantage already disadvantaged groups, such as low-SES and ethnic minority students. Test-optional admission policies are seen as one way to diversify the student body. However, research on the effects of test-optional admissions on inequalities is inconclusive (e.g., Belasco, Rosinger, and Hearn 2015; Bennett 2022; Saboe and Terrizzi 2019). This research mainly focuses on the role of test performance but overlooks differences in test participation and their potential effect on college admission.

Little research focuses on the effects of test optionality for gender differences, perhaps because women, who are typically researched as a disadvantaged group, constitute the majority of higher education students today. Yet research on gender differences is theoretically and policy-wise of interest. In general, test optionality might increase (and not reduce) differences in test performance and test participation between applicant groups—but the pattern for SES and gender might differ: In contrast to SES or ethnicity, where differences in test participation are likely to increase the majority group's advantage, test optionality might counteract women's higher GPAs and actually increase gender diversity in college admissions. Moreover, female applicants' higher GPAs, on average, allow us to distinguish between

compositional and behavioral explanations for differences in test participation.

In this study, we therefore examined gender inequality in test-based admissions and the role of test participation versus performance for the male advantage in a competitive, high-stakes setting. To this end, we exploit the unique German context of test-optional admissions to highly selective medical schools. In contrast to existing studies, the German context offers a unique opportunity for disentangling the two mechanisms—test participation and performance—because the alternative to test-optional is not test-mandatory (as in the United States and other countries) but test-free, so test avoidance is a real option for applicants.

Our study reveals a male advantage in test performance (*ceteris paribus* of GPA)—but only among test-takers with poorer grades, not among test-takers with an excellent or good GPA. Concerning test participation, we find gender differences only when controlling for GPA (i.e., between male and female applicants with similar GPAs)—suggesting women's lower test-participation rate is not due to their better average GPAs and thus their lower need to compensate with test scores. Behavioral rather than compositional (GPA differences) explanations thus seem to be at work. Furthermore, both male and female applicants with a medium GPA show higher test-participation rates, thus supporting our need to compensate—most gains thesis (i.e., a nonlinear, inverse U-shaped relationship between GPA and test participation). Moreover, we find gender differences in test participation only among applicants with a medium GPA.

Applicants in the test-optional context can also choose not to report test scores despite taking the test. Our additional analyses (based on 2018 survey data) show that applicants indeed use this option in the German context. Gender differences in the share of nonreporters and the difference between reported and achieved test scores are overall too small to change our main findings on gender differences in test-based admissions programs. In summary, our findings suggest that gender differences in test participation and test performance might be caused by gender differences in self-assessment and competitiveness, test anxiety, and stereotype threat. We cannot study these mechanisms directly with the data at hand. Thus, further research is needed in this respect.

Our analyses also show that gender differences in test participation and test performance increase the admission chances of male applicants, with

male applicants' better test performance being the main source of their higher admission chances (*ceteris paribus* of GPA). In terms of the magnitude of this male advantage, it is important to note that the female advantage in GPA—the selection criterion with the highest weight in admissions to German medical programs—and the male advantage in test performance neutralize each other, leading to almost equal admission chances for male and female applicants.

What are the policy implications of our findings for test-optional admissions? The German context, with its mix of test-optional and test-free admissions, provides an ideal environment for students to pursue their test-taking preferences without having to sacrifice applying for a specific field or institution. We find that even in this free-choice setting, the male advantage in test-optional admissions is mainly caused by differences in test performance. This means that in both test-optional and test-mandatory settings, male applicants benefit from advantages in test performance (and less so from advantages in test participation). This similarity in findings under different test conditions also demonstrates that only under test-free conditions (no admission tests at all), the male advantage in test-taking and test-performance vanishes. Yet in terms of gender inequality, a corresponding request would be to establish admission policies that also eliminate the female advantage in GPA.

In contrast to gender, for other social stratification categories like SES or ethnic groups, admission tests—no matter whether they are optional or mandatory—are likely to increase inequality in admissions because, in this case, tests are unlikely to compensate for poorer GPAs of the disadvantaged groups. Privileged socioeconomic groups would still be more likely to use this low-risk situation (because poor test scores are not penalized) to their advantage, drawing on their economic, cultural, and information resources to better prepare for the tests and improve their outcomes (see Mbekeani 2023). Thus, for these stratification categories, a test-free system or ensuring a level playing field in terms of test preparation would enhance social diversity in the student body.

Finally, although test-based admissions increase gender diversity in the student body in female-dominated fields, our study also demonstrates the

limited potential of admission tests in this respect. The key drivers of the skewed gender composition of medical students are not admissions or admission tests but the much lower number of male applicants (only 35 percent in Germany, see Table 1). Thus, the main source of gender disparities lies in gender differences in occupational aspirations. These differences might be caused by the higher prevalence of the caregiving motive in female occupational aspirations (Barone 2011) and the ample opportunities for boys to realize their income aspirations via other fields of study (like STEM, law, or economics). Boys may also be more discouraged from studying medicine in anticipation of the required GPA.

ACKNOWLEDGMENTS

We thank our student assistants (Julia Bersch, Birte Freer, Anna Hetz, Johanne Klindworth, Carlotta Rieble, and Patrick Hölzgen) for their manifold support and the Stiftung für Hochschulzulassung for data access. We received valuable feedback from Rourke O'Brien, David Brady, and participants of the colloquium of the WZB research unit "Skill Formation and Labor Markets." An earlier version of this article was presented at the 2021 RC28 meeting in Turku.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from the German Science Foundation (Grant No. SO 430/13-1).

RESEARCH ETHICS

The results presented in this article do not allow for deductive disclosure of applicants' identities. This research received ethical approval from the WZB research ethics committee (No. 2018/1/32).

ORCID iD

Claudia Finger  <https://orcid.org/0000-0002-3424-7998>

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

NOTES

1. For an overview of the number of test-takers between 1986 and 2018, see https://fairtest.org/sites/default/files/ACT-SAT-Annual-Test-Takers-Chart_0.pdf.
2. The very few private universities offering medical programs are not part of the central admission system and are thus not included in our data.
3. The Center for Higher Education publishes the most prominent (field-specific) ranking of German universities every three years based on administrative information and evaluations by students and professors. Our analyses of these data for medical programs (based on the years 2012, 2015, and 2018) show low variation between TMS-based and test-free programs on mean values of publication per professor (16.1 vs. 15.1) or citation per publication (6.4 vs. 5.5)—as indicators of research reputation—or in teaching quality as assessed by students (2.0 vs. 2.3 on a scale from 1 = high to 6 = low).
4. Our observation period stops in 2018 because a major reform of the central admission system was implemented afterward, making it unfeasible to add further years to answer our research questions. A decision of the Federal Constitutional Court in 2017 generated changes in the admission procedure, including a further increase in test-optional admissions for medical programs. The start in 2012 is driven by data availability.
5. We did not include types of university entrance certificate because only 4 percent of our applicants obtained a specialized (instead of a general) higher education entrance certificate.
6. The share of applicants with a GPA of 3.0 to 4.0 who applied via the waiting quota is 88 percent, compared to only 29 percent among those with a GPA of 1.0 to 1.1.
7. This increase in the likelihood of gaining admission is calculated as follows: Table 4 shows that an increase in test scores of 1 SD raises the likelihood of admission by 15.5 percentage points; 1 SD in test scores is 0.53 (see Table D2 in the online Supplemental Material). Hence, 0.163 (Model 1) amounts to 0.31 SD, and this multiplied by 15.5 results in 4.8 percentage points. Similar calculations apply to 0.183 (in Model 2), resulting in 5.4 percentage points. The relative increase is calculated using the overall admission likelihood in our study sample of test-takers: 4.8/31.4 and 5.4/31.4, respectively (see Table D3 in the online Supplemental Material).
8. This difference is statistically significant at the 0.01 level, based on the test for “seemingly unrelated estimates” (*suest* command in Stata), which is appropriate for the comparison of nested models.
9. The data are available online: 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). doi:10.7802/2515.

REFERENCES

- Alon, Sigal, and Marta Tienda. 2007. “Diversity, Opportunity, and the Shifting Meritocracy in Higher Education.” *American Sociological Review* 72(4): 487–511.
- Azen, Razia, Shmuel Bronner, and Naomi Gafni. 2002. “Examination of Gender Bias in University Admissions.” *Applied Measurement in Education* 15(1): 75–94.
- Barone, Carlo. 2011. “Some Things Never Change: Gender Segregation in Higher Education across Eight Nations and Three Decades.” *Sociology of Education* 84(2):157–76.
- Belasco, Andrew S., Kelly O. Rosinger, and James C. Hearn. 2015. “The Test-Optional Movement at America’s Selective Liberal Arts Colleges.” *Educational Evaluation and Policy Analysis* 37(2):206–23.
- Bennett, Christopher T. 2022. “Untested Admissions: Examining Changes in Application Behaviors and Student Demographics under Test-Optional Policies.” *American Educational Research Journal* 59(1):180–216.
- Berggren, Caroline. 2007. “Broadening Recruitment to Higher Education through the Admission System.” *Studies in Higher Education* 32(1):97–116.
- Booth, Alison, and Patrick Nolen. 2012. “Choosing to Compete: How Different Are Girls and Boys?” *Journal of Economic Behavior & Organization* 81(2):542–55.
- Boulis, Ann K., and Jerry A. Jacobs. 2008. *The Changing Face of Medicine*. Ithaca, NY: Cornell University Press.
- Breen, Richard, Kristian B. Karlson, and Anders Holm. 2018. “Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models.” *Annual Review of Sociology* 44:39–54.
- Buchmann, Claudia, Dennis J. Condron, and Vincent J. Roscigno. 2010. “Shadow Education, American Style: Test Preparation, the SAT and College Enrollment.” *Social Forces* 89(2):435–61.
- Buchmann, Claudia, Thomas A. DiPrete, and Anne McDaniel. 2008. “Gender Inequalities in Education.” *Annual Review of Sociology* 34:319–37.
- Cahlíková, Jana, Lubomír Cingl, and Ian Lively. 2020. “How Stress Affects Performance and Competitiveness across Gender.” *Advances in Management Science* 66(8):3295–310.
- Camara, Wayne J., and Krista Mattern. 2022. “Inflection Point: The Role of Testing in Admissions Decisions in a Postpandemic Environment.” *Educational Measurement: Issues and Practice* 41(1):10–15.
- Chapell, Mark S., Z. Benjamin Blanding, Michael E. Silverstein, Masami Takahashi, Brian Newman, Aaron

- Gubi, and Nicole McCann. 2005. "Test Anxiety and Academic Performance in Undergraduate and Graduate Students." *Journal of Educational Psychology* 97(2):268–74.
- Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106(6):1691–730.
- Correll, Shelley J. 2004. "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations." *American Sociological Review* 69(1):93–113.
- DiPrete, Thomas A., and Claudia Buchmann. 2013. *The Rise of Women*. New York, NY: Russell Sage Foundation.
- Dunleavy, Dana M., Marc H. Kroopnick, Keith W. Dowd, Cynthia A. Searcy, and Xiaohui Zhao. 2013. "The Predictive Validity of the MCAT Exam in Relation to Academic Performance through Medical School." *Academic Medicine* 88(5):666–71.
- Finger, Claudia. 2022. "(Mis)Matched College Aspirations and Expectations: The Role of Social Background and Admission Barriers." *European Sociological Review* 38(3):472–92.
- Furuta, Jared. 2017. "Rationalization and Student/School Personhood in U.S. College Admissions: The Rise of Test-Optional Policies, 1987 to 2015." *Sociology of Education* 90(3):236–54.
- Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics* 118(3):1049–74.
- Grodsky, Eric, John R. Warren, and Erika Felts. 2008. "Testing and Social Stratification in American Education." *Annual Review of Sociology* 34(1):385–404.
- Jackson, Michelle, Tatiana Khavenson, and Tatiana Chirkina. 2020. "Raising the Stakes: Inequality and Testing in the Russian Education System." *Social Forces* 98(4):1613–35.
- Jurajda, Štěpán, and Daniel Münich. 2011. "Gender Gap in Performance under Competitive Pressure: Admissions to Czech Universities." *The American Economic Review* 101(3):514–18.
- Kadmon, Guni, and Martina Kadmon. 2016. "Academic Performance of Students with the Highest and Mediocre School-Leaving Grades: Does the Aptitude Test for Medical Studies (TMS) Balance Their Prognoses?" *GMS Zeitschrift für medizinische Ausbildung* 33(1):Doc7.
- Klasik, Daniel. 2012. "The College Application Gauntlet: A Systematic Analysis of the Steps to Four-Year College Enrollment." *Research in Higher Education* 53:506–49.
- Kozu, Tadahiko. 2006. "Medical Education in Japan." *Academic Medicine* 81(12):1069–75.
- Leiner, Julia E. M., Thomas Scherndl, and Tuulia M. Ortner. 2018. "How Do Men and Women Perceive a High-Stakes Test Situation?" *Frontiers in Psychology* 9:2216. doi:10.3389/fpsyg.2018.02216.
- Lovaglia, Michael J., Jeffrey W. Lucas, Jeffrey A. Houser, Shane R. Thye, and Barry Markovsky. 1998. "Status Processes and Mental Ability Test Scores." *American Journal of Sociology* 104(1):195–228.
- Mau, Wei-Cheng, and Richard Lynn. 2001. "Gender Differences on the Scholastic Aptitude Test, the American College Test and College Grades." *Educational Psychology* 21(2):133–36.
- Mayer, Karl U., Walter Müller, and Reinhard Pollak. 2007. "Germany: Institutional Change and Inequalities of Access in Higher Education." Pp. 240–65 in *Stratification in Higher Education*, edited by R. Arum, A. Gamoran, and Y. Shavit. Stanford, CA: Stanford University Press.
- Mbekeani, Preeya Pandya. 2023. "Income-Based Gaps in College-Going Activities: High School Classes of 1992 and 2004." *Sociology of Education* 96(1):62–79.
- Nankervis, Bryan. 2011. "Gender Inequities in University Admission Due to the Differential Validity of the SAT." *Journal of College Admission* 213:24–30.
- Niederle, Muriel, and Lise Vesterlund. 2011. "Gender and Competition." *Annual Review of Economics* 3:601–30.
- Ors, Evren, Frédéric Palomino, and Eloiç Peyrache. 2013. "Performance Gender Gap: Does Competition Matter?" *Journal of Labor Economics* 31(3):443–99.
- Penner, Andrew M., and Robb Willer. 2019. "Men's Overpersistence and the Gender Gap in Science and Mathematics." *Socius* 5(3). doi:10.1177/2378023118821836.
- Powell, Justin J. W., and Heike Solga. 2011. "Why Are Participation Rates in Higher Education in Germany so Low?" *Journal of Education and Work* 24(1):49–68.
- Protivinský, Tomáš, and Daniel Münich. 2018. "Gender Bias in Teachers' Grading." *Studies in Educational Evaluation* 59:141–49.
- Puddey, Ian B., and Annette Mercer. 2014. "Predicting Academic Outcomes in an Australian Graduate Entry Medical Programme." *BMC Medical Education* 14(1):31. doi:10.1186/1472-6920-14-31.
- Rosinger, Kelly O., Karly Sarita Ford, and Junghee Choi. 2021. "The Role of Selective College Admissions Criteria in Interrupting or Reproducing Racial and Economic Inequities." *The Journal of Higher Education* 92(1):31–55.
- Rubin, Allen. 1985. "Significance Testing with Population Data." *Social Service Review* 59(3):518–20.
- Saboe, Matt, and Sabrina Terrizzi. 2019. "SAT Optional Policies: Do They Influence Graduate Quality, Selectivity or Diversity?" *Economics Letters* 174:13–17.
- Saygin, Perihan O. 2020. "Gender Bias in Standardized Tests: Evidence from a Centralized College Admissions System." *Empirical Economics* 59:1037–65.

- Shurchkov, Olga. 2012. "Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints." *Journal of the European Economic Association* 10(5):1189–213.
- Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. "Stereotype Threat and Women's Math Performance." *Journal of Experimental Social Psychology* 35(1):4–28.
- Stumpf, Heinrich, and Douglas N. Jackson. 1994. "Gender-Related Differences in Cognitive Abilities: Evidence from a Medical School Admissions Testing Program." *Personality and Individual Differences* 17(3):335–44.
- Summit Educational Group. 2021. "College Admission Testing Guide." <https://mytutor.com/college-admission-testing-guide/>.
- Tiffin, Paul A., John C. McLachlan, Lisa Webster, and Sandra Nicholson. 2014. "Comparison of the Sensitivity of the UKCAT and A Levels to Sociodemographic Characteristics." *BMC Medical Education* 14(1):7. doi:10.1186/1472-6920-14-7.
- Voyer, Daniel, and Susan D. Voyer. 2014. "Gender Differences in Scholastic Achievement: A Meta-analysis." *Psychological Bulletin* 140(4): 1174–204.
- Zeidner, Moshe. 1990. "Does Test Anxiety Bias Scholastic Aptitude Test Performance by Gender and Sociocultural Group?" *Journal of Personality Assessment* 55(1–2):145–60.
- Zhang, Yu, and Mun Tsang. 2015. "Gender Gap in the National College Entrance Exam Performance in China: A Case Study of a Typical Chinese Municipality." *Asia Pacific Education Review* 16(1):27–36.
- Zwick, Rebecca. 2019. "Assessment in American Higher Education: The Role of Admissions Tests." *The ANNALS of the American Academy of Political and Social Science* 683(1):130–48.

Author Biographies

Claudia Finger is a research fellow in the research unit "Skill Formation and Labor Markets" at the WZB - Berlin Social Science Center. Her research focuses on social background and gender inequality in (higher) education. Her work has been published in the *European Sociological Review*, *Sociology of Education*, *Social Science Research*, and *Research in Social Stratification and Mobility*.

Heike Solga is director of the research unit "Skill Formation and Labor Markets" at the WZB - Berlin Social Science Center and professor of sociology at the Freie Universität Berlin. Her research interests are education and labor market research from a life course perspective. She is one of the principal investigators of the German National Education Panel Study (NEPS), responsible for the school-to-work transition period.