

Evaluation Criteria and Women's Attainment of Elite STEM Education: Evidence from College Admission Records

Wei-hsin Yu,^a Kuo-Hsien Su^b

a) University of California, Los Angeles; b) National Taiwan University

Abstract: Research on women's underrepresentation in science, technology, engineering, and mathematics (STEM) fields rarely addresses the roles of institutional gatekeepers and their screening criteria. Using full application records of the most prestigious university in Taiwan, we examine how the assessment criteria used by departments to determine admissions shape women's relative chance of entering elite STEM programs. Results from department fixed-effect models indicate that male-dominated STEM programs actually rate female applicants' written application materials and interviews higher. Female applicants are still less likely admitted to such programs than males because many STEM departments also use major-specific tests, which are not strictly curriculum based and impose great competitive pressure on selected students. Even the highest-achieving female students with a strong STEM interest perform worse than males in this type of tests, especially when the tests are given by male-dominated departments. Because of this gender performance gap, female students' chances of being admitted to elite STEM programs continue to be obstructed even as the college admission system became holistic and incorporated assessment criteria that could favor females.

Keywords: gender inequality; STEM education; evaluation methods; elite college admission; test performance

Reproducibility Package: The authors received special permission to use the confidential data of applications of National Taiwan University (NTU) for this publication and are prohibited from sharing the data. Access to the NTU application data should be requested directly from the Office of Admission under NTU's Office of Academic Affairs (<https://www.aca.ntu.edu.tw/w/acaEN/Contact>). However, all of the code files and ancillary data generated from publicly available sources are stored in Dataverse and can be obtained through <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/KUVUS8>.

Citation: Yu, Wei-hsin, Kuo-Hsien Su. 2025. "Evaluation Criteria and Women's Attainment of Elite STEM Education: Evidence from College Admission Records" *Sociological Science* 12: 357-387.

Received: February 24, 2025

Accepted: May 1, 2025

Published: June 23, 2025

Editor(s): Ari Adut, Elizabeth Bruch

DOI: 10.15195/v12.a16

Copyright: © 2025 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 

GIrls and women all over the world have been making headways in educational attainment, to the extent that many industrialized countries now observe a gender gap favoring women in college attendance and completion (Buchmann, DiPrete, and McDaniel 2008). Despite this progress, women continue to be underrepresented in many science, technology, engineering, and mathematics (STEM) fields in higher education (Gerber and Cheung 2008; Charles and Bradley 2009), especially within elite universities (Bielby et al. 2014). This underrepresentation contributes to men's dominance in STEM-related occupations and ultimately reinforces the belief in male superiority in STEM areas, which further dissuades the next generation of girls from pursuing science education (Brotman and Moore 2008; Stearns et al. 2016).

Much research has examined the reasons behind female underrepresentation in STEM education, with most focusing on factors discouraging girls' engagement in science or their interest in STEM majors in college (Brotman and Moore 2008; Gerber and Cheung 2008; Charles and Bradley 2009; Legewie and DiPrete 2014; Xie, Fang and Shauman 2015; Raabe, Boda, and Stadtfeld 2019). Although some studies have indicated unequal treatments between boys and girls by science and mathematics teachers (e.g., Li 1999), few consider the roles of institutional gatekeepers in higher education. STEM majors, especially in highly selective universities, often have high requirements for people interested in choosing them. In many parts of the world where students must be admitted to a department at the same time of college acceptance (Yu 2009; Bordon and Fu 2015; Machado and Szerman 2021; Ma et al. 2023), the barriers set by STEM departments for admission can affect women's chances of becoming top scientists or engineers even more directly. Nevertheless, we know surprisingly little about how institutional gatekeepers and their selection criteria shape the gender gap in admissions to STEM programs, especially those in elite universities across the world.

In this study, we use a unique data set containing key information of all applications submitted in 2022 and 2023 to the departments across National Taiwan University (NTU), the most prestigious and competitive university in Taiwan (Luoh 2002; Yu and Su 2024), to examine how admission criteria and gender performance gaps by assessment type contribute to the unequal representation between men and women in elite STEM programs. Similar to other East Asian countries (Lee and Park 2014; Kariya 2020), Taiwan began educational reforms in the 2000s and shifted its college admission system, which had been entirely based on a once-a-year standardized entrance exam, to a more holistic one, considering not only such an exam but also high school and extracurricular performances, written statements, and evaluations through interviews and major-specific tests (Liu 2022; Yu and Su 2024). As the Taiwanese system requires students to be admitted by the department, not just the university, of their choice, the shift to a holistic admission system encouraged the diversification of selection methods across college departments, with some placing a greater emphasis on tests, whereas others on written materials or interviews. The variation in NTU departments' admission criteria, along with their evaluation scores of applicants' written application materials, interviews, and major-specific tests in our data set, provides a unique opportunity to shed light on assessment criteria's roles in shaping the admission chances of girls and women who already aspire to enter elite STEM programs.

We specifically ask: (1) Does the gender performance gap vary by the assessment type and field of study? (2) Does the shift to a holistic admission system, which incorporates a variety of assessment criteria, alter overall gender representation in STEM programs at elite colleges? (3) How do the screening methods used by elite STEM programs, especially male-concentrated programs, affect women's relative admission chances compared to men's? The case of NTU is useful for addressing these questions because it is indisputably the highest ranking and most desirable university in Taiwan (Luoh 2002; Global Views Magazine 2024; Times Higher Education 2024), to which girls of the greatest STEM aptitude and achievement in the country would apply. The highly selective nature of NTU applicants largely reduces

gender differences in academic aspirations and orientations, making any unequal outcomes between male and female applicants likely attributable to the gatekeepers' evaluation methods. The prestige and competitiveness of NTU also give us the rare opportunity to provide systematic evidence on elite colleges' admission processes, which are typically opaque (Karabel 2005).

Beyond elite college admissions, our analysis of gender differences in performances under different types of screening methods across departments adds to the literature on gender and evaluation formats and designs in the educational system (Hamilton 1998; Buchmann et al. 2008; Niederle and Vesterlund 2010). Prior research primarily focuses on how girls' and boys' academic achievements differ when assessed through grades versus test scores (e.g., Downey and Vogt Yuan 2005; Graetz and Karimi 2022). Our study expands the existing literature by showing how selecting students through national standardized tests, written application materials, interviews, and specialized major-specific tests can each give advantage to males or females, and by uncovering whether the gender advantage may also depend on the admitting department's characteristics (e.g., STEM vs. others). Altogether, this study provides an important understanding of how institutional gatekeepers and men's and women's performance differences by assessment type jointly shape gender representation in STEM fields.

Gender and Stem Majors in College

Numerous studies have documented gender segregation across fields of study in higher education (Gerber and Cheung 2008). Women worldwide are far less likely to major in many STEM fields, particularly those pertaining to math, computer science, physical science, and engineering, than men (Charles and Bradley 2009; Ramirez and Kwak 2015; Riegle-Crumb and Peng 2021). Because having a STEM major in college is a key predictor of entering STEM occupations (Sassler et al. 2017), and because such occupations tend to pay more, the gender gap in STEM education directly contributes to a lower lifetime income for women (Haveman and Beresford 2012; Sullivan et al. 2018).

Much research on women's underrepresentation in STEM majors emphasizes factors shaping women's choices of such majors. Although girls have no worse math and science grades or lower STEM course participation than boys in high school (Xie et al. 2015), their belief in male superiority, self-perception of math and science ability, gendered self-images and career plans, weaker emphasis on earning prospects in considering majors, greater tendency to avoid competition, and lack of peer support for STEM pursuits are all thought to be responsible for gender disparities across fields of study in college (Correll 2004; Charles and Bradley 2009; Niederle and Vesterlund 2010; Raabe et al. 2019; Brenøe and Zölitz 2020; Quadlin 2020; Weeden, Gelbgiser, and Morgan 2020; Riegle-Crumb and Peng 2021). Altogether, these explanations provide a thorough understanding of the shortage of female supply in STEM fields. In contrast, the demand side of the story is understudied. Even when researchers highlight high school or college contexts, their focus is still on how environmental characteristics cultivate or dampen female students' interest in STEM (Griffith 2010; Legewie and DiPrete 2014; Xie et al. 2015).

However, do women with sufficient interest and motivation always have equal access to college STEM majors as do men?

Few researchers have examined gender inequality in admission chances among those already aspiring to enter STEM programs. Although a small number of qualitative studies of Ph.D. programs' admission practices have shed some light on this issue (Posselt 2015; Posselt, Porter, and Kamimura 2018), a systematic comparison of women's and men's access to a variety of STEM programs is still absent. Filling this void is especially critical for many parts of the world where the student quotas for STEM departments are relatively fixed and the process of being admitted into one is extremely competitive. Throughout East Asia and in various Latin American and European countries, students must be simultaneously admitted to a department and the college in which it is nested, with each department having a predetermined number of spots available (Yu 2009; Bordon and Fu 2015; Machado and Szerman 2021; Ma et al. 2023). Under this system, female and male students alike must compete to enter STEM programs, especially the ones in elite colleges. The ways in which prestigious STEM departments select aspiring students, and how their selection methods may facilitate gender inequality, are therefore worthy of attention. The case of elite STEM program admission is particularly useful because gender differences in STEM aspirations and preparations are likely insignificant among the extremely high achieving who aim to enter such programs (O'Dea et al. 2018); thus, we can largely rule out supply side explanations. Understanding factors shaping women's representation in elite STEM departments is also important in its own right, as these departments tend to produce top scientists and prominent role models (Zimmerman 2019), who can inspire the next generation of girls (Brotman and Moore 2008).

Evaluation Methods and Gendered Performance

Institutional gatekeepers and their admission criteria could play a role in the gender representation in elite STEM programs because male and female students may perform differently according to the assessment type and study subject (Hamilton 1998; Downey and Vogt Yuan 2005; Graetz and Karimi 2022). Research shows that girls tend to score lower than boys in math and science tests, especially beyond middle school and on advanced topics (Niederle and Vesterlund 2010; Kahn and Ginther 2018; Bahar 2021), but they often outperform boys in reading tests (Marks 2008) and earn better grades in both math and reading (Downey and Vogt Yuan 2005; Buchmann et al. 2008). Many argue that these performance differences are rooted in the gender system and associated cultural beliefs (Ridgeway and Correll 2004; Guiso et al. 2008; Kahn and Ginther 2018). Because math and science are viewed as "masculine" domains in many societies (Brandell and Staberg 2008), the cultural belief in men's superior task competence tends to be more salient and consequential when girls and boys are evaluated for math and science achievements (Correll 2001). Although this belief can generally lead girls to underestimate their ability in math and science (Correll 2004) and develop anxiety that negatively affects their achievement performance (Devine et al. 2012), it tends to manifest more in test scores than grades. This is because girls are also socialized to be more cooperative

and better at exerting effortful control over impulsive or unproductive behavior than boys (Hyde 2014). The better classroom participation, homework effort, and work ethic can compensate girls' competence belief, facilitating their relatively high math and science grades, but the same traits do not reward them as much in tests (Downey and Vogt Yuan 2005).

Even among math and science tests, the gender performance gap is small or nonexistent for in-class or low-stake tests (Felson and Trudeau 1991; Hyde and Linn 2006); the gap mainly exists in high-stake standardized math tests such as the scholastic assessment test (SAT) (Niederle and Vesterlund 2010; Bahar 2021; Graetz and Karimi 2022), which are often a key criterion for college admissions. Drawing from Kimball's (1989) familiarity theory, Felson and Trudeau (1991) contend that girls are better at tests that are clearly curriculum based and have well-defined boundaries and contents—that is, tests based on familiar materials—than tests that require applications of novel concepts or learning from outside of the classroom, as in the case of high-stake standardized tests. Supporting this argument, Hamilton (1998) finds that even within a science test, girls' accuracy rates are especially lower than boys' in items that gauge knowledge that is not strictly conveyed in the classroom.

Another theory for the especially wide gender performance gap in high-stake standardized tests is that women are socialized to be warm and supportive, making them less comfortable than men with competitions. Laboratory experiments show that women perform worse than men in tests given for the sake of competing for resources or limited spots but not tests without competitive pressure (Niederle and Vesterlund 2010; Gërkhani, Brandts, and Schram 2023). As exams designed for college admission purposes ultimately require test takers to be engaged in a zero-sum competition, girls perform relatively worse in such exams than in-class tests.

Both the familiarity and competition theories would predict that an admission system that downplays performance in high-stake standardized tests will benefit women. A couple of recent studies indeed show that a shift to a test-optional policy, or simply reducing the weight of standardized test scores in college admission, is associated with an increase in female share among admitted students (Bennett 2022; Yu and Su 2024), but counterevidence, that the SAT-optional movement has no effect on the student body's gender composition, also exists (Saboe and Terrizzi 2019). The mixed findings may be partly due to the variability in how much colleges with test-optional admissions still factor in test scores and partly related to the other evaluation criteria introduced into the process when college admission becomes more holistic. In a holistic admission system, universities and departments can differ widely in whether beside national standardized exam scores, they require extensive written materials (e.g., high school transcripts, personal statements, extracurricular records, and writing samples), interviews, or college- or department-specific tests, and in how they weigh the various types of assessments in the final admission decisions (Bastedo 2021). Despite much scholarly interest in how the use of SAT scores affects inequality in college admission (Alon and Tienda 2007; Grodsky, Warren, and Felts 2008), we know relatively little about gender-related advantages or disadvantages tied to other assessment methods. We also do not know whether

the assessment-specific gender performance gap is conditional on the field of study. To answer how admission systems contribute to the female share in elite STEM programs, we need to learn which other evaluation methods are used in lieu of standardized tests and whether such methods introduce new gender disparities.

Unlike national standardized exams, which tend to hide test-takers' gender, assessments through interviews, essays, or other written materials are subjected to evaluators' conscious or unconscious gender bias. According to expectation states theory, gender as a status characteristic is conducive to the belief in greater male competence, especially in regard to socially prescribed masculine tasks, which results in others using more lenient standards to judge men than women (Correll 2004; Correll and Ridgeway 2006). This theory would expect the double standards to disadvantage females when male-dominated STEM programs assess applicants through written materials or interviews. At the same time, because gender socialization encourages girls to be more communicative (Leaper and Friedman 2007), they tend to exhibit better writing skills than boys (Petersen 2018). It is possible that the female writing advantage could compensate the harm caused by evaluators' use of stricter standards in assessing females' essays and other written materials, although real-world evidence on such evaluations is scant. Nevertheless, experimental research finds that science faculty at highly selective colleges rate women's application materials lower when considering for a laboratory job (Moss-Racusin et al. 2012).

The applicants' gender is even more salient in face-to-face interviews than in written applications, making the gender status belief potentially more impactful (Correll and Ridgeway 2006; Ridgeway 2013). The fact that the gender gap in verbal skills is small also means that girls may not be able to make up the disadvantage caused by this belief in interviews (Hyde 2014; Petersen 2018). Supporting the status-based disadvantage, research shows that in male-dominated work settings, such as prestigious orchestras, hiding gender identities in auditions increases women's likelihood of receiving an offer (Goldin and Rouse 2000). Following the same logic, female applicants may be rated worse than males in STEM-field interviews in which applicants cannot hide their gender. However, a few case studies of medical schools' or medicine residency's admission processes find that the interview scores do not depend on the applicant's gender (Hamidi et al. 2021; Stern et al. 2021). Still, it is unclear whether this absence of gender effect will be found in fields with far larger shares of male students, such as engineering and physical science, than medicine.

In many East Asian countries and elsewhere, universities and departments also conduct their own major-specific tests and consider such test performance in admission decisions (Bastedo 2021; Wang, Tang, and Bao 2024). Compared to national standardized tests, for which prior-year exam questions and private test-preparation services are readily available, specialized tests designed by college departments to gauge students' aptitude or fit for the specific discipline are even less closely tied to high school curriculums and harder to prepare (too small scaled for private agencies to offer test-preparation services, among others). Based on the familiarity theory, female students may be especially likely to perform worse than males in this type of tests. Because such tests are usually given after college applicants pass the initial screening (Yu and Su 2024), the zero-sum competitive

nature of the tests is particularly salient, which could further disadvantage female test takers (Gërkhani et al. 2023).

In sum, prior research predominantly focuses on gender performance differences in grades versus tests, especially high-stake standardized tests (Downey and Vogt Yuan 2005; Buchmann et al. 2008; Grodsky et al. 2008). We still lack knowledge on how other assessment methods that are increasingly introduced into college admission processes worldwide may affect gender inequality in admission chances. Because gender as a status characteristic is especially relevant in male-concentrated STEM fields (Correll 2004), the gender performance gap may vary by not only the assessment type but also the admitting department's area of study, making elite STEM programs' admission criteria relevant to women's relative chances of entering them.

The Context of Taiwan

Similar to other East Asian countries, higher education in Taiwan has undergone rapid expansion and change in recent decades (Hannum et al. 2019). Currently, Taiwan's compulsory education is 12 years, with nearly all high school graduates advancing to college. The competition for acceptance to a top-ranking college nonetheless remains fierce, as the public's high consensus and keen awareness of college ranking make attending a prestigious university exceptionally desirable (Hannum et al. 2019; Yu and Su 2024). Along with the college expansion, women's attendance to four-year universities has also increased. Figure 1 shows that by the mid-2010s, women's share in Taiwanese colleges has surpassed men's. At NTU, however, the number of women has continued to be behind men's. The proportion female has been even lower in NTU's STEM programs; only slightly over one-third of STEM students were women in 2023.

Figure 1 also shows that the female share of college students and that of NTU students began to diverge around 2000, which, interestingly, was also when Taiwan's college admission reforms started. Before the reforms, college admissions were exclusively based on the results of a once-a-year national college entrance examination, while students can now utilize multiple channels to obtain a spot in college (Liu 2022). Currently, the most popular channel is through direct applications to college departments (Yu and Su 2024). More similar to the admission to U.S. graduate than undergraduate programs, Taiwanese students must be simultaneously admitted to a department and the university to which it belongs. Each department in Taiwanese universities has a predetermined number of seats irrespective of the volume of applications.¹ All Taiwanese students applying for colleges have to first take a national standardized test, the general scholastic ability test (GSAT), which is offered once a year. College departments conduct a first-round screening solely based on the GSAT score, selecting a number of applications, which is twice or three times of the available seats, into the second round. Although all programs use GSAT scores for the first-round screening, they have considerable autonomy in deciding what to consider among the GSAT test subjects. Departments can choose to only consider results from certain subjects in the GSAT (e.g., advanced math,

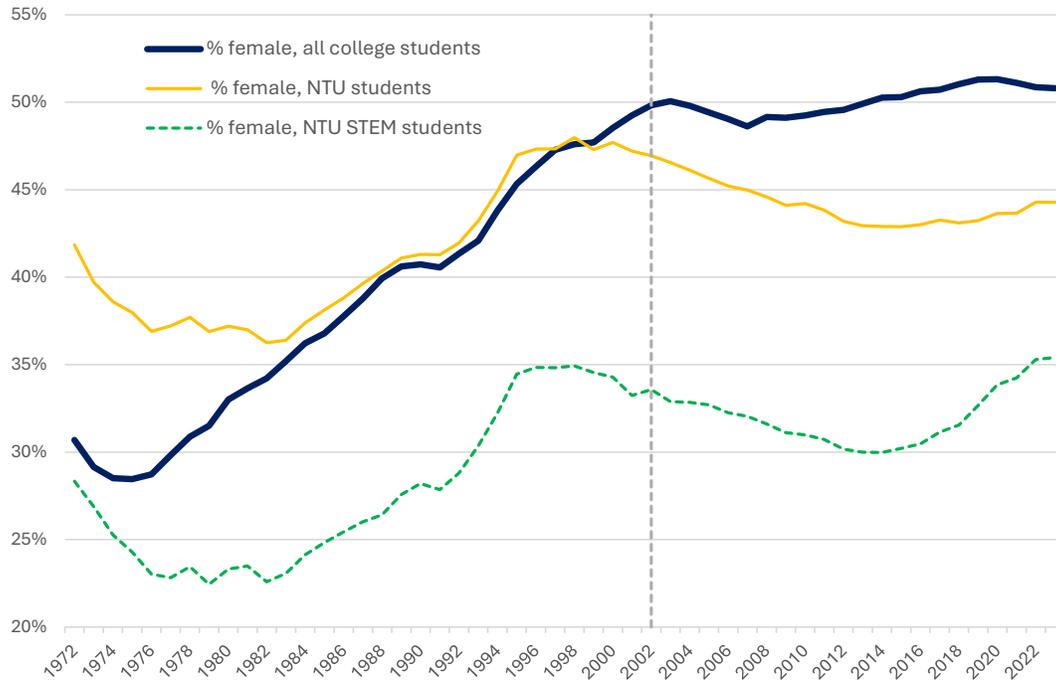


Figure 1: Changes in female enrollment at Taiwanese colleges and NTU, 1972–2023. Source: Ministry of Education, Republic of China (Taiwan), various years. *Note:* The gray vertical line indicates the year when the multichannel college entrance system was formally implemented. Individuals enrolled in NTU’s College of Engineering, College of Science, College of Public Health, College of Life Science, College of Bio-Resources and Agriculture, College of Medicine, and College of Electrical Engineering and Computer Science are considered as STEM students.

English, natural science only; no regular math, Chinese, or social science) or weigh the results of some subjects more than the others.

At the time of application, high school seniors (or graduates) also submit an application dossier, which includes the applicant’s high school transcript, personal statements, records of extracurricular merits, and examples of prior academic work such as a science experiment report or a term paper. Department faculty only read and score the dossiers of applicants that have entered the second round. Aside from the dossiers, departments can decide which other evaluation criteria they would use to assess second-round candidates. Some programs also conduct personal interviews, whereas others require applicants to take a major-specific test (or tests), which could be in the form of writing proofs for high-level math problems, producing computer codes to solve a specific issue, or answering essay questions to demonstrate advanced knowledge in Chinese literature. It is also possible to require both interviews and major-specific tests. After the second-round evaluations, departments then use their specific formula to calculate each application’s final score, giving the four assessment components—GSAT, dossier, interview, and specialized test scores—varying weights in their formula. The final admission decisions are entirely based on the ranking of applicants’ final scores, with departments accepting the top-scoring applications until all seats are filled. To

ensure transparency, all college departments must advertise their first- and second-round selection criteria, including how they weigh the GSAT subjects in the first round, the requirements of an interview, a specialized test, or both in the second round, and the specific formula they use to calculate the final score, at the beginning of the admission season.

The substantial variation in college departments' selection criteria enables us to examine not only whether the shift to a more holistic admission system affects gender inequality in acceptance chances to elite STEM programs but also how the adoption of each assessment method and the weight placed on it influence these chances. With a unique data set containing applicants' scores from standardized exams, dossier evaluations, interviews, and specialized tests, we can precisely show gender performance differences by evaluation method and admitting department's STEM status or male dominance, thus adding to the broader literature on assessment criteria and educational inequality. Because using department-specific criteria to admit students is not unique to Taiwan and is common at the postgraduate level in many countries (Bastedo 2021; Wang et al. 2024), our study can also have implications for the role of institutional gatekeepers in women's pursuits of STEM education elsewhere.

Data and Methods

Data

This study uses data from NTU's application records in 2022 and 2023. In the two years together, NTU received 30,355 applications from 16,192 students, with some applying to multiple departments at once. Among the applications, 11.4 percent were eventually accepted, although this number is not strictly comparable with acceptance rates of universities in Western countries. Because Taiwanese universities' use of GSAT scores for the first-round screening makes it easy for students to assess their chances of passing that round (Liu 2022), NTU is likely to receive far fewer applications from ambitious but not competitive applicants than other elite colleges in the West,² resulting in its seemingly higher acceptance rate. The selection to apply to NTU during the years examined was likely based on the competitiveness of students' GSAT scores, rather than their economic or other considerations, given that NTU's tuition is low (about \$1000 per semester), its location is superior, and the Confucianism's legacy has led students from all class backgrounds to greatly value acceptance to a high-ranking college (Li and Xie 2020).

Our application records include the department to which the application was submitted and the applicant's gender (male or female, no other category available), high school, high school grade point average (GPA), high school graduation year, permanent home address, and GSAT results by subject. We also have the scores of all second-round evaluations and the final admission outcome. Because NTU admissions are conditional on an application selected into the second round, an application would have no dossier, interview, or major-specific test scores if it did not pass the first round. Therefore, we limited our analysis to all applications

in the second round. During the two years, 8,478 applications entered the second round.³ A handful of applications received 0 or ineligible scores for the follow-up interview or major-specific test (3.8 percent), suggesting that the applicants were either extremely ill-prepared or did not show up for the interview or test. To prevent outliers driving the results, we excluded these observations in the part of the analysis where we separately examine dossier scores, interview scores, and major-specific test scores. Our additional analysis indicated that including the cases with a "0" score did not affect the results in any meaningful way.

Because the digital records kept by NTU only include applicants' high school GPA when their high school is able to electronically link their transcript to the application, the GPA information is missing for those not enrolled in a Taiwanese high school at the time of application.⁴ This group includes individuals who graduated from a Taiwanese high school in an earlier year or attended high school in other countries. We used multiple imputations to fill in GPA information for these cases, creating 20 imputed data sets. Because our regression models control for graduating in an earlier year, GSAT results, and high school ranking, all of which are very closely related to the high school GPA, using imputed GPA for missing cases should not affect our results much. In fact, when we excluded the GPA in an exploratory analysis, the results were virtually unchanged.

Analytic Strategy and Measurement

Our analysis begins by examining how the gender performance gap varies by assessment type and whether this gap depends on the characteristics of the department assessing applications. We estimated department fixed-effect models predicting scores from three types of evaluations: dossiers, interviews, and major-specific tests. The use of fixed-effect models enables us to account for time-invariant unobserved program characteristics, such as the program's overall culture and general appeal to female (or male) students. The inclusion of department fixed effects also transforms the analysis into comparisons within the same departments. The dossier and interview were both evaluated by department faculty and scored on a 0–100 scale. The major-specific test scores are more complicated, because some programs require multiple tests with uneven weights, whereas others require just one. We calculated the weighted score from multiple tests, if more than one test was given, and then standardized all departments' total specialized test scores to a 0–100 scale. For all three outcomes, we first examined the gender gap in scores, after accounting for extensive sociodemographic differences among applicants (details below). We then introduced interactions between the applicant's gender and the department's characteristics, including its STEM status, male-student share, and male-faculty proportion, to show whether the gender performance gap is especially large in STEM or male-dominated programs.

In the analysis, we included the department's STEM status, male-student share, and male-faculty share in separate models because they are fairly correlated, despite some conceptual distinction among them (e.g., not all STEM programs are male dominant). We measured whether a program is in a STEM field based on the college to which it belongs. NTU has 11 colleges. We considered departments in

the College of Engineering, College of Science, College of Public Health, College of Life Science, College of Bio-Resources and Agriculture, College of Medicine, and College of Electrical Engineering and Computer Science as STEM ones. To specifically focus on male-dominated programs, we created an additional indicator of the department's proportion of male students in the application year, using public records of student composition at NTU. We also used public records to construct the share of male faculty in the department, as aspiring students' perception of women friendliness of a discipline could be based on the faculty gender ratio. Moreover, male faculty members may discriminate female applicants more than female faculty members in interviews. As shown in Figure S1 in the online supplement, most STEM departments at NTU have majority-male student or faculty bodies, whereas few departments in non-STEM colleges have more male than female students (Economics and Information Management being notable exceptions). About 22 percent of NTU's departments, all in STEM fields, had three quarters or more of their students being male. Over three quarters of the departments had more male than female faculty members, because relatively few women successfully obtained faculty positions at NTU even in fields popular with female students. The shares of male faculty and students are generally correlated (0.74 in the analytic sample).

The next part of our analysis addresses the question whether the incorporation of the additional evaluation methods, on the whole, alters the gender composition of NTU's STEM or male-dominated programs. We specifically used the GSAT information included in each application and the final admission outcome to create a counterfactual analysis. Because GSAT, as a national standardized test aiming to gauge students' high school learning, is similar to the national college entrance exam used before the admission reforms, and because the number of seats in each department is predetermined, we were able to rank applications based on the GSAT scores and identify those that would have been accepted if based on a national standardized test alone.⁵ We then compared these applicants with the ones actually admitted in the current system, which takes into account other assessment results, to show whether the shift to a holistic system changes the same applications' outcomes.

Previous research using natural experiments, such as the switch to a test-optional policy, to investigate the effects of admission standards generally faces the difficulty discerning the treatment from its unintended impact on the applicant pool (e.g., Bennett 2022). For example, it is possible that racial diversity increases with the move to a test-optional policy because more students from minority groups apply, rather than such students actually fare better (Belasco, Rosinger, and Hearn 2015). In contrast, our counterfactual analysis identifies the "switchers," whose admission outcome would have been different between the exam-only and holistic systems, and "non-switchers," whose outcomes are unchanged either way, within the *same* group of applicants. Because the non-switchers, who are always admitted or always rejected, would not contribute to any shifts in student composition across NTU departments, we focus on the switchers to assess how the change in the admission system may have affected women's shares in male-dominated STEM departments. We specifically fitted linear probability models with department fixed effects among the switchers to predict a positive (from rejected to admitted) instead of negative

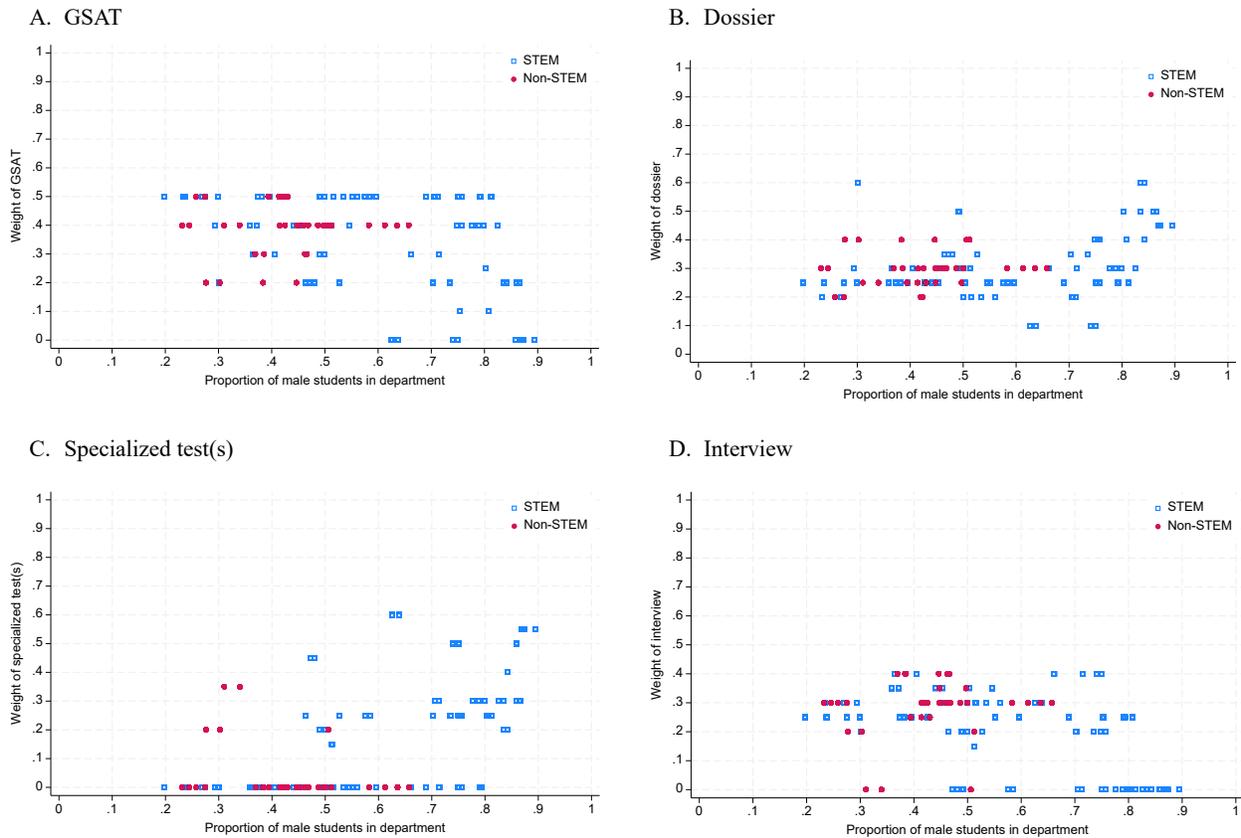


Figure 2: Weights of admission criteria by department's proportion of male students and STEM category. *Note:* Each mark in the figures represents a department year. There are two years of data for each department.

(from admitted to rejected) change with the move from the exam-only to current admission systems. We also included interactions between the applicant's gender and the department characteristics (i.e., STEM status, male-student share, and male-faculty share). If, after considering many observable applicant traits, gender remains predictive of a positive rather than negative change, and the gender effect is significant in male-dominated STEM departments, then we can argue that the shift to a holistic system likely has resulted in a net increase or decrease in female students in such departments.

Finally, to answer which exact components of departments' evaluation criteria explain their potentially different gender gaps in admission chances, we estimated department fixed-effect models predicting the probability of admission among all second-round applications. We specifically analyzed how the program-specific gender admission gap changes after accounting for the program's assessment requirements and assigned weights to the requirements. As mentioned before, NTU programs assign different weights to the GSAT, dossier, interview, and major-specific test scores to calculate the final ranking of applications. Figure 2 shows these differences by the departments' proportion of male students and STEM status. Table S1 in the online supplement further lists each department's second-round

criteria, along with the lowest standardized GSAT score for entering the second round, by admission year. In general, programs in STEM fields, especially those with high proportions of male students, are more likely to require one or more specialized tests in the second round, although their weights assigned to the test results are similar to non-STEM departments', if the latter require such a test. We created dummy variables indicating whether the target department requires an interview and whether it requires a major-specific test. We also measured the weights each department assigned to the dossier, interview, major-specific test, and GSAT scores, respectively.

Control variables. All our multivariate models include applicants' socioeconomic conditions and other characteristics that may affect their prior learning opportunities and hence evaluation scores. The Taiwanese government considers students from remote islands (e.g., Kinmen) and low-income households as having disadvantaged status; hence, we constructed a dummy variable reflecting this status. Next, we used each applicant's permanent home address to approximate their family wealth, given that the application records contain no information about students' family income or parental occupations. We linked the address to the publicly available residential home sale records to calculate the median sales price per *ping* (a local measuring unit, $\sim 3.31 \text{ m}^2$) from the 20 most recent transactions on the same street or town (if rural areas) and approximated the applicant's family home worth using this value. Taiwan's home ownership rate is very high (>85 percent), and the permanent home is generally the main source of residents' wealth (Chen 2020). The housing price also tends to vary greatly between different neighborhoods within the same city or across localities but very little within the same block. For these reasons, our estimated per-*ping* cost of applicants' permanent home is a reasonable proxy for the wealth of their family of origin. We took a natural logarithm of the housing price (in the unit of 1,000 New Taiwanese dollars) to adjust for skewness. The third variable to gauge the applicant's socioeconomic background is the proportion of college graduates among those aged 25–29 in their permanent home's zip code area. Neighborhoods where college graduates concentrate are likely to have more educational resources for residents.

We also controlled for the applicant's high school ranking and whether the high school is public or private (private = 1). Because the public in Taiwan tend to rank high schools based on their college placement results—for example, more admissions to NTU—we used the applicant's high school's proportion of students passing NTU's first-round screening to rank high schools. Our data include the actual names of the high school, which allows us to obtain each school's total number of graduates and calculate the percentage entering NTU's second round. In addition, we created a binary indicator for the applicants who graduated from high school before the application year, because such applicants probably had applied for college in an earlier year (i.e., their senior year) and failed to be admitted to their desired program. The earlier graduation year could send a signal of inferior quality and thus have a negative effect on the application assessment.

To account for the applicant's academic knowledge and skills, which could explain their assessment results, we also included the applicant's high school GPA (on a 0–100 scale). Because the application dossiers contain applicants' high school

transcripts, having GPA in the models predicting dossier scores means that the results reflect faculty evaluations of the rest of the dossier contents, including applicants' essays, prior academic works, and reports of extracurricular merits. Except for the counterfactual analysis in which GSAT scores are already factored in when constructing the dependent variable, we also included the GSAT results to further approximate the applicant's ability to perform in academic settings. We standardized the GSAT scores across departments to scores on a 0–100 scale to enable comparisons.⁶ Although the NTU administrative data we obtained do not contain the actual essays or writing samples in applicants' dossiers, they recorded the dossier length. Because there is no page limit to the dossier length, those submitting longer ones are likely to have included more prior scholarly works and activity records,⁷ which also implies greater academic effort. Therefore, we used the dossier page length as a proxy of both the quality of the applicant's prior records and the effort they put into the application.

We further included the number of NTU departments to which the applicant submitted applications. Individuals can apply to up to five departments at once. The number of applications submitted could indicate the applicant's application effort, risk-taking tendency, and level of interest in the targeted program, all of which may affect the admission outcome. In the models, we introduced five dummies for having submitted 1–5 applications, respectively. Finally, we took into account the application year, given that our data are from two years.

Table S2 in the online supplement shows descriptive statistics of the analytic sample by applicant gender. The two groups are fairly similar in socioeconomic backgrounds. About 69 percent and 70 percent of the applications from female and male students, respectively, would have received the same admission result in an exam-only admission system as in the current holistic system. This is not surprising because high-achieving students tend to excel, whereas lower-achieving students perform relatively poorly, regardless of evaluation methods. The fact that as much as 30 percent of the applications could have different outcomes, however, suggests that evaluation methods used in admission systems are consequential.

Results

Descriptive Findings

Figure 3 shows key gender differences between applications to STEM and non-STEM programs, with a positive value indicating a gender gap favoring women and a negative value favoring men. As shown in Figure S2 in the online supplement, the results are similar if narrowing the STEM programs to those with high proportions of male students (>60 percent). Among non-STEM applications, female applicants have higher scores in all performance categories and longer dossiers. In contrast, female applicants to STEM programs, on average, have lower GSAT and major-specific test scores than males,⁸ but their GPA values, dossier lengths, and interview scores are still higher. Interestingly, female students aspiring to enter STEM programs applied to significantly fewer NTU programs than their male counterparts. This gender gap remains when focusing on applications to

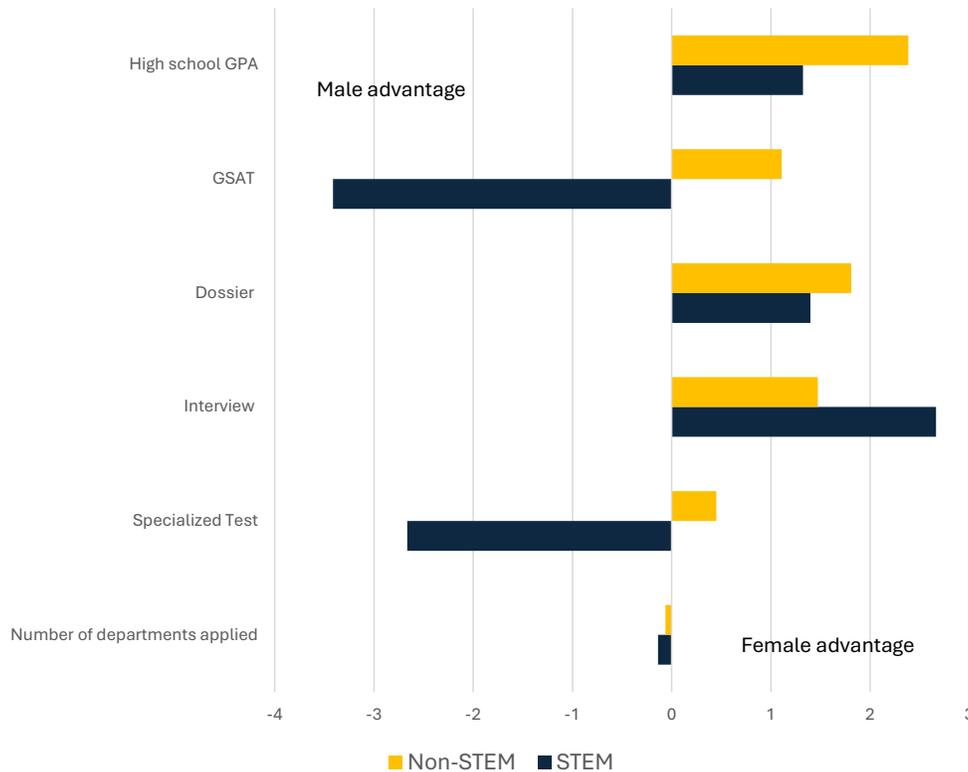


Figure 3: Gender differences between applications to STEM and non-STEM departments. *Note:* The gender differences are calculated as the average of female applications minus the average of male applications, thus a negative number indicating the value is higher for men's applications. All gender differences are statistically significant at the 0.05 alpha level except for the total number of departments applied among applications to non-STEM programs. Among STEM applications, 1,674 are from female students and 3,680 from male students. Among non-STEM applications, 1,792 are from female applicants and 1,332 from male applicants.

male-dominated STEM programs only (Fig. S2). These findings suggest that female STEM-program applicants are more focused and likely more interested in their chosen majors than male ones. As the masculine image of male-dominated STEM fields likely dissuades women from pursuing related education, it is no surprise that women choosing to apply to male-concentrated STEM programs appear especially selective and determined.

Gender Performance Gaps by Evaluation Method

To show how the gender performance gap varies by assessment method and the field of study, Tables 1–3 show main results from department fixed-effect models predicting the scores of application dossiers, interviews, and major-specific tests, respectively (full model results in Tables S3–S5 in the online supplement). Model 1 in Table 1 demonstrates a within-department female advantage in dossier scores. This advantage remains after accounting for high school ranking, GPA, dossier length, and standardized GSAT result, all of which approximate applicants' academic

Table 1: Department fixed-effect models predicting dossier scores.

	Model 1	Model 2	Model 3	Model 4	Model 5
Female applicant	2.476** (0.163)	0.471** (0.121)	0.245 (0.172)	-0.376 (0.477)	-0.949* (0.410)
Female applicant × STEM program			0.389 (0.228)		
Proportion male faculty in department				-1.567 (2.760)	
Female applicant × proportion male faculty				1.246 (0.657)	
Proportion male students in department					3.989 (3.661)
Female applicant × proportion male students					2.610** (0.727)
Other applicant/application characteristics	No	Yes	Yes	Yes	Yes
Intercept	81.681** (0.134)	25.795** (1.837)	25.876** (1.838)	26.943** (2.754)	23.630** (2.769)
Number of observations	8396	8396	8396	8396	8396

Note: Model results presented here are based on estimations from 20 imputed data sets. All models include department fixed effects and are clustered by applicants. Other applicant/application characteristics include the applicant's disadvantaged status, graduation status, log family home worth, residential area's proportion of college graduates, private school attendance, high school grade average, dossier page length, GSAT score, number of applications submitted to NTU, and application year. Numbers in parentheses are clustered standard errors.

** $p < 0.01$; * $p < 0.05$.

abilities and efforts (Model 2). As the application dossier contains GPA information, our finding of a female advantage net of GPA levels suggests that female applicants receive higher scores for their essays and writing samples.⁹ Models 3 and 4 indicate that programs in STEM fields or with higher proportions of male faculty are not more discriminatory against female applicants; if anything, evaluators in such programs seem to score female applicants' dossiers higher. Model 5 shows that the gender performance gap varies with the admitting department's male student share. The results, shown in Figure S3 in the online supplement, indicate that female applications receive higher dossier scores than male applications in departments with 37 percent or more male students, and the female advantage increases as the departmental share of male students grows.

Table 2 presents similar models for interview scores. The analytic sample for these models is smaller because not all programs required an interview in the second round. Model 1 shows a gender gap favoring females in interview performance, which can be partially explained by indicators of applicant backgrounds and abilities (Model 2). This female advantage is marginally greater in STEM departments (Model 3) and is more pronounced in departments with higher proportions of male faculty (Model 4). Model 5 further indicates that departments with greater shares of male students do not rate female applicants' interviews significantly differently. Thus, contrary to the expectation of status characteristics theory, male faculty or

Table 2: Department fixed-effect models predicting interview scores.

	Model 1	Model 2	Model 3	Model 4	Model 5
Female applicant	2.175** (0.212)	1.059** (0.202)	0.767** (0.273)	-0.310 (0.615)	-0.100 (0.732)
Female applicant × STEM program			0.641 (0.379)		
Proportion male faculty in department				10.627** (3.905)	
Female applicant × proportion male faculty in dept.				2.148* (0.940)	
Proportion male students in department					-10.075 (5.376)
Female applicant × proportion male students in dept.					2.397 (1.484)
Other applicant/application traits	No	Yes	Yes	Yes	Yes
Intercept	83.046** (0.185)	52.754** (2.574)	52.903** (2.567)	46.061** (3.642)	57.763** (3.693)
Number of observations	5249	5249	5249	5249	5249

Note: Model results presented here are based on estimations from 20 imputed data sets. All models include department fixed effects and are clustered by applicants. Other applicant/application characteristics include the applicant's disadvantaged status, graduation status, log family home worth, residential area's proportion of college graduates, private school attendance, high school grade average, dossier page length, GSAT score, number of applications submitted to NTU, and application year. Numbers in parentheses are clustered standard errors.

***p* < 0.01; **p* < 0.05.

Table 3: Department fixed-effect models predicting major-specific test scores.

	Model 1	Model 2	Model 3	Model 4	Model 5
Female applicant	-3.224** (0.607)	-4.489** (0.586)	-0.808 (0.958)	2.499 (1.937)	4.266* (1.682)
Female applicant × STEM program			-4.180** (1.147)		
Proportion male faculty in department				0.882 (15.537)	
Female applicant × proportion male faculty in dept.				-8.879** (2.606)	
Proportion male students in department					25.725 (24.514)
Female applicant × proportion male students in dept.					-12.803** (2.565)
Other applicant/application characteristics	No	Yes	Yes	Yes	Yes
Intercept	65.196 (0.404)	-20.535** (6.927)	-21.699** (6.929)	-22.217 (14.568)	-40.542* (19.367)
Number of observations	3713	3713	3713	3713	3713

Note: Model results presented here are based on estimations from 20 imputed data sets. All models include department fixed effects and are clustered by applicants. Other applicant/application characteristics include the applicant's disadvantaged status, graduation status, log family home worth, residential area's proportion of college graduates, private school attendance, high school grade average, dossier page length, GSAT score, number of applications submitted to NTU, and application year. Numbers in parentheses are clustered standard errors.

***p* < 0.01; **p* < 0.05.

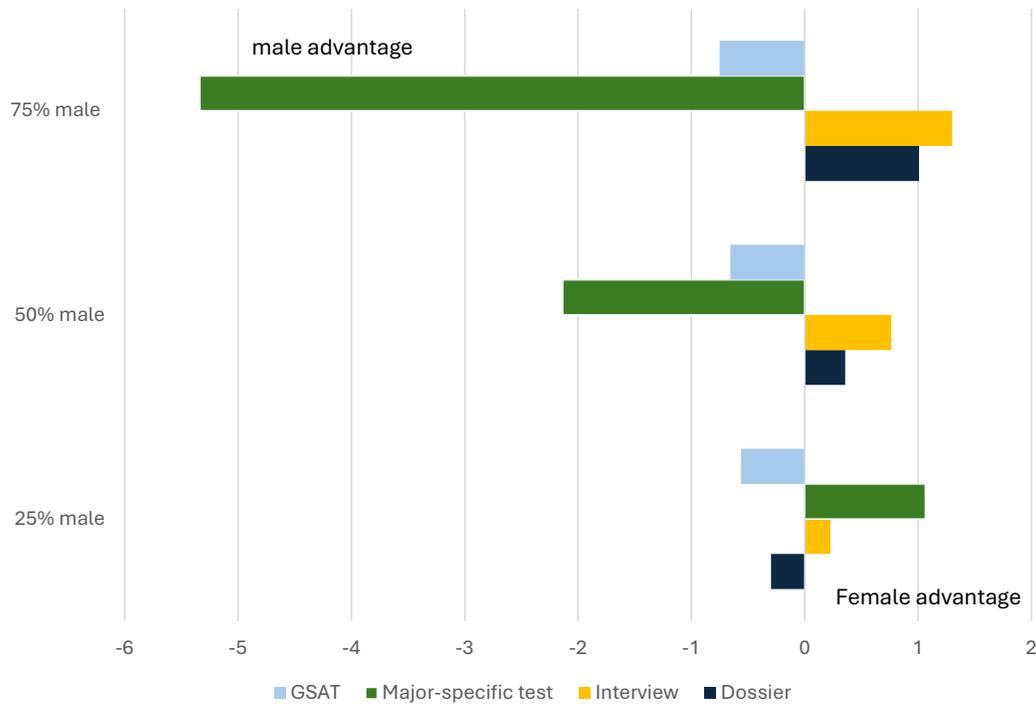


Figure 4: Predicted gender differences in second-round evaluation results. *Note:* The labels “75 percent male,” “50 percent male,” and “25 percent male” indicate that the department’s male student share is 75 percent, 50 percent, and 25 percent, respectively, except for the predicted interview scores, for which they indicate the department’s male faculty share. The predicted gender differences are calculated from the female coefficients and the interactions with department male share, using the results from Model 5 in Table 2, Model 4 in Table 3, Model 5 in Table 4, and Model 5 in Table S7.

faculty in programs with high proportions of male students show no tendency to discriminate against female applicants in interviews.

Table 3 presents department fixed-effect models predicting the standardized total score from all major-specific tests taken during the second round. The models include the same variables as in the previous two tables, but the analytic sample is yet smaller because a number of departments did not require a major-specific test. Unlike for dossier or interview scores, female applicants have lower major-specific test scores, even after controlling for GSAT and high school GPA scores, which reflect their knowledge in STEM subjects (Model 2). In fact, the female disadvantage increases from Model 1 to Model 2, indicating that the gender performance gap would have been wider if not for the fact that female applicants tend to be better students (i.e., with higher GPA levels) and from more advantageous backgrounds than males applying to the same program. The gender gap is more pronounced with STEM programs (Model 3), programs with higher proportions of male faculty (Model 4), and programs with larger shares of male students (Model 5). As shown in Figure S3 in the online supplement, only in departments with less than 33 percent male students do female students perform no worse than males in department-specific specialized tests.

Table 4: Department fixed-effect models predicting positive instead of negative changes in admission outcomes.

	Model 1	Model 2	Model 3	Model 4	Model 5
Female applicant	0.183** (0.026)	0.061* (0.025)	0.163** (0.038)	0.155* (0.076)	0.338** (0.076)
Female applicant × STEM program			-0.167** (0.047)		
Proportion male faculty in department				-0.120 (0.669)	
Female applicant × proportion male faculty in dept.				-0.137 (0.105)	
Proportion male students in department					-0.137 (0.830)
Female applicant × proportion male students in dept.					-0.501** (0.130)
Other applicant/application characteristics	No	Yes	Yes	Yes	Yes
Intercept	0.424** (0.016)	-1.984** (0.256)	-1.964** (0.256)	-1.914** (0.545)	-1.942** (0.554)
Number of observations	2608	2608	2608	2608	2608

Note: Model results presented here are based on estimations from 20 imputed data sets. All models include department fixed effects and are clustered by applicants. Other applicant/application characteristics include the applicant's disadvantaged status, graduation status, log family home worth, residential area's proportion of college graduates, private school attendance, high school grade average, dossier page length, GSAT score, number of applications submitted to NTU, and application year. Numbers in parentheses are clustered standard errors.

** $p < 0.01$; * $p < 0.05$.

Figure 4 summarizes the results from Tables 1 to 3, with the predicted gender gaps by evaluation method and the target department's level of male dominance. For a comparison, we also include the predicted gender differences in GSAT scores, using Model 5 in Table S7 in the online supplement. We used the proportion of male students to indicate the department's male dominance in all cases except for when we calculated the predicted interview scores, for which we used the proportion of male faculty because it corresponds to the interview scores more (Table 2). Using student gender ratio for all outcomes nonetheless yielded similar results. The figure indicates that female applicants mostly have advantages in dossier and interview scores, and these advantages are greater in more male-dominated departments. However, male applicants enjoy net advantages in major-specific test scores. In a department where 75 percent of the students are male, the predicted male advantage in such scores is far greater than the female advantages in dossier or interview scores, and it considerably exceeds any male advantage in GSAT scores.

Holistic Admission System and the Counterfactual

The analysis thus far has shown that the additional evaluations in Taiwan's current admission system benefit female applicants in some ways. However, has the overall shift to a holistic admission system increased women's share in male-concentrated STEM programs in elite colleges? Table 4 addresses this question with a series of department fixed-effect models predicting a positive rather than negative change in

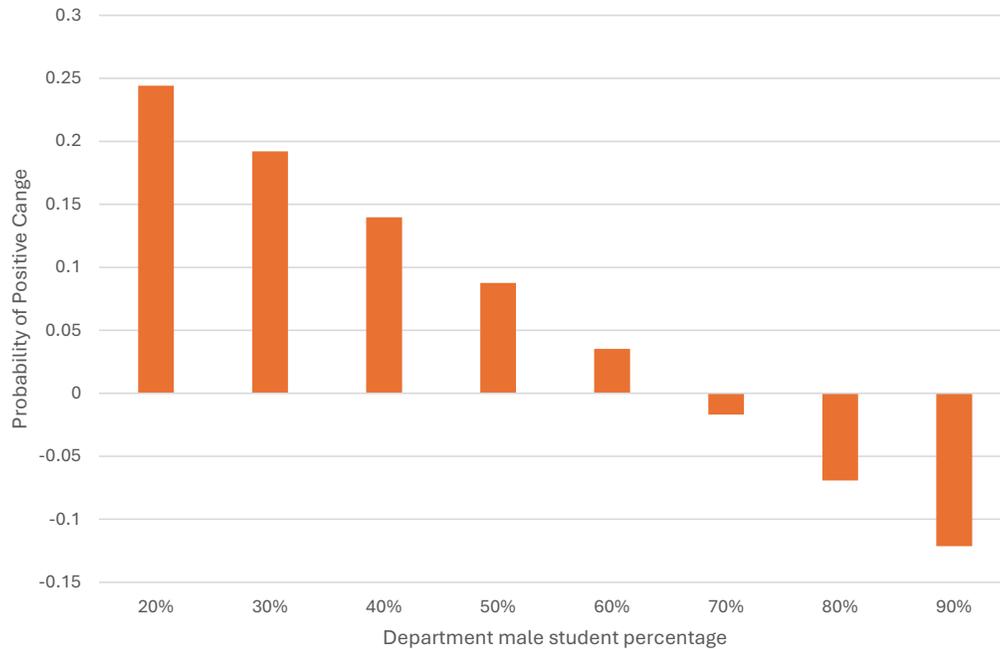


Figure 5: Changes in female advantage with the switch to holistic system by department gender composition. *Note:* Each bar indicates the probability for female applicants undergoing a positive than negative change in the admission outcome subtracting the probability of male applicants, holding other factors constant. Values are calculated from Model 5 in Table 5.

the admission outcome from an exam-only system to the current one (full model results in Table S8 in the online supplement). Model 1 shows that female applications are more likely among those undergoing a positive than negative change, suggesting a rise in the female student share at NTU with the shift to the current system. The female advantage diminishes considerably in Model 2, which adds applicant characteristics. Our additional analysis, not presented here, revealed that female applicants' higher GPA and, to a less extent, longer dossier length explain most of the change in female advantage from Models 1 to 2. However, the gender effect is uneven across departments. Model 3 indicates that women's applications are more likely to be among those experiencing positive than negative switches for non-STEM programs, whereas there is no net gender difference among applications to STEM programs ($0.163-0.167 \sim 0$). The female advantage with the change in the admission system does not vary by the department's male faculty share (Model 4), but it clearly corresponds to the gender composition among students (Model 5). Figure 5 shows the interaction results in Model 5 and indicates that in programs with 68 percent or more male students, female applicants' advantage would ultimately disappear or reverse.

Taken together, results from our counterfactual analysis suggest that the shift to a holistic admission system has affected the student gender ratio across elite college departments unevenly. Female applicants gain no advantage in admissions to NTU's male-dominated STEM programs with the system change, even though compared to male applicants, they receive higher dossier scores in departments

with high proportions of male students, and they have higher interview scores in departments with high concentrations of male faculty. The uneven results across departments underscore the need to examine how exactly institutions differ in their assessment methods as holistic college admissions become increasingly prevalent worldwide. In the next part of analysis, we turn to differences in NTU departments' admission criteria and how they account for the gender gap in admission chances.

Explaining Variation in the Gender Admission Gap

Table 5 presents a series of department fixed-effect models predicting the acceptance of an application, with the same control variables as those in Tables 1–3. Model 1 is the baseline model, showing that female applicants' relative probability of admission decreases with an increase in the male-student share in the department. We focus on the variation with the share of male students because Table 4 suggests that the extent of female advantage in the current admission system depends more on this share than the department's faculty gender composition, and this share likely captures variation across departments better than the binary measure of STEM status. Model 2 indicates that female applicants' chance of admission decreases when the department requires any major-specific tests. This requirement ultimately explains why women's relative probability of admission declines in programs with higher proportions of male students, as the coefficient for the interaction between female applicant and department's proportion of male students reduces considerably from Model 1 to Model 2.

Model 3 includes the weights placed on the different assessment types by the department. The weights of application dossiers or interviews are not relevant to women's relative chances of admission. The interaction between female applicants and the weight on major-specific test(s) is negative and significant, but Model 4 suggests that this coefficient is driven by the wide gap in women's admission probabilities between the departments giving any weights to major-specific test results and those giving no weight (i.e., not requiring a test). According to Model 4, placing a heavy weight (>30 percent) on major-specific tests does not disadvantage women's admission chances more than placing a moderate weight (= 30 percent). The results were similar when we used different categories for the weight (e.g., 0 percent, 1 percent to 24 percent, and 25 percent and more). Thus, even when departments weigh the major-specific tests little in their final ranking of applications, the admission chances of women's applications still diminish; just having the requirement of major-specific tests is sufficient to dampen female applicants' admission chances.

Model 5 in Table 5 investigates whether the effect of requiring a specialized test for female applicants depends on how male dominated the evaluating department is. The results show that having the test requirement is especially detrimental to female applicants when the department giving the test has a higher proportion of male students. Conversely, female applicants' admissions to programs with higher proportions of male students are not significantly affected when the program requires no specialized tests. It is possible that female test takers especially underestimate their own competence and hence underperform when facing a non-

Table 5: Department fixed-effect models predicting the probability of admission.

	Model 1	Model 2	Model 3	Model 4	Model 5
Female applicant	0.130** (0.039)	0.093 (0.060)	0.114 (0.079)	0.075 (0.042)	0.033 (0.046)
Proportion male students in department	-1.187** (0.420)	-1.218** (0.422)	-1.268** (0.422)	-1.255** (0.420)	-1.279** (0.421)
Female × proportion male students	-0.233** (0.067)	-0.072 (0.083)	-0.084 (0.085)	-0.064 (0.080)	
Interview required		-0.107 (0.141)			
Female × interview required		-0.015 (0.035)			
Major-specific test(s) required		0.011 (0.119)			
Female × major-specific test(s) required		-0.111** (0.034)			0.078 (0.064)
Interview weight in final decision			0.551 (0.620)		
Female × interview weight			0.018 (0.117)		
Major-specific test weight in final decision			1.096* (0.547)		
Female × major-specific test weight			-0.224* (0.101)		
Dossier weight in final decision			0.224 (0.228)		
Female × dossier weight			-0.145 (0.143)		
Weight of major-specific test(s) (<i>ref.</i> 0%):					
<30%				0.092 (0.065)	
≥30%				0.215** (0.082)	
Female × major-specific test weight <30%				-0.116** (0.031)	
Female × major-specific test weight ≥ 30%				-0.071 (0.046)	
Female × major test required × proportion male students					-0.151* (0.069)
Female × no major test × proportion male students					0.025 (0.093)
Number of observations	8478	8478	8478	8478	8478

Note: Model results presented here are based on estimations from 20 imputed data sets. The models include department fixed effects and are clustered by applicants. Numbers in parentheses are clustered standard errors. All models control for disadvantaged status, graduation status, log family housing price, concentration of college graduates in residential area, high school ranking, high school being private, high school grade average, dossier length, and the number of the applications submitted to NTU. Numbers in parentheses are clustered standard errors.

** $p < 0.01$; * $p < 0.05$.

curriculum-based high-competitive-pressure test in a customarily male field. As a result, we find a male-favoring gender admission gap only among programs that are male dominated *and* require a major-specific test.

Discussion and Conclusions

Research on gender representation in STEM fields rarely examines the role of institutional gatekeepers and their evaluation standards on women's chances of entering elite STEM programs. Using application records from the most prestigious university in Taiwan, we have shown that a reduced emphasis on national standardized tests in college admissions, on average, benefits female applicants, which is consistent with the findings of a few recent studies (Bennett 2022; Yu and Su 2024). However, the female advantage has not extended to male-concentrated STEM departments. The reason for the differences across departments is that the various screening criteria introduced into the holistic college admission system do not have the same gender implications, and their implications often depend on the evaluating department's STEM status and level of male dominance. Thus, similar to what previous research shows for hiring (Uhlmann and Cohen 2005), gatekeepers can maintain the status quo by shifting the evaluation criteria even as they face social and cultural changes that potentially favor women.

On the gender implications of specific assessment methods, we have demonstrated that contrary to status characteristics theory's expectation for gender bias to manifest in STEM faculty's evaluations of application dossiers, faculty members of programs with high proportions of male students, typically in STEM fields, tend to score female applicants' dossiers higher than those of less male-concentrated programs do. This female advantage remains even after controlling for many applicant characteristics that may affect evaluators' impression, such as the applicant's high school GPA, high school ranking, GSAT score, and dossier length. The analysis of interview results similarly offers no support that male-dominated STEM programs rate female interviewees lower. If anything, programs with large shares of male faculty, which are related to but not necessarily in STEM fields, are likely to give higher interview scores to female applicants. These results are somewhat consistent with gender socialization theory, which anticipates women's advantages in writing and communication, although the theory alone cannot explain why the advantages are observed more among applications to male-concentrated STEM departments. Perhaps female students aspiring to enter heavily male-dominated STEM programs are also highly selective; their exceptionally strong determination and intellectual ability show in their dossier contents and interviews, leading to especially higher scores.

However, with respect to the performance on department-specific specialized tests, we have found a clear gender gap favoring males. This gap cannot be explained by indicators of applicants' effort, skills, and knowledge at all. The gender gap also widens as the department's share of male students increases. The male advantage in major-specific test performance offsets any female advantage in dossier or interview scores and ultimately lowers female applicants' likelihood to be admitted to male-concentrated STEM programs at NTU.

Our finding that many NTU departments' reliance on specialized tests virtually negates or reverses any advantages female applicants have in college admission highlights the important roles of program gatekeepers and the prevailing assessment standards in STEM fields in shaping women's chances of becoming top scientists. Not only do girls and women fare worse when assessed by standardized tests but they also face constrained educational opportunities when the gatekeepers and disciplines in general believe in the validity of tests that are not based on a well-defined curriculum. Unlike prior research concerning gender differences in responses to various test formats and contents (e.g., Hamilton 1998; Coffman and Klinowski 2020), which typically examines these differences among general test takers, our study provides exceptionally strong evidence on the effect of assessment type by comparing a selected group of male and female students. Because female applicants with the ambition and desire to attend NTU's heavily male concentrated and highly prestigious STEM programs could be more selective than their male counterparts—their higher dossier scores indeed suggest so—their worse major-specific test scores are unlikely a result of their truly inferior ability or interest in the specific area; rather, it is likely to reflect females' comparatively worse performance in exams that have highly uncertain contents and impose heightened competitive pressure.

Our analysis suggests that cultural beliefs likely play a role as well, as the requirement of major-specific tests hampers female students' admission chances more in departments with higher proportions of male students. The more male dominated a field of study is, the more likely women and men alike will believe that men have superior ability in that field. When the test given by the field has no well-defined boundaries and study materials, the belief in male superiority becomes particularly critical; it could amplify female test takers' but reduce male test-takers' anxiety toward the uncertainty, resulting in a greater gender gap in test performance.

As much as our results support the importance of cultural beliefs, we also find that women's admission chances to elite STEM programs, even male-dominated ones, would not be obstructed had the programs not required a specialized, non-curriculum-based test. The negative effect of specialized tests on female access to NTU's STEM departments exists regardless of whether the department places much or little weight on the test score in its final admission decisions. These findings have general implications because even in a context like the United States, where colleges' major-selection processes differ from Taiwan's, similar specialized tests are often used in the school system to sort students into different tracks or programs (e.g., magnet vs. regular programs). For example, students entering new schools or at a certain transition point often have to take math or science placement tests, which may not be based on a specific curriculum with well-defined boundaries. Just similar to the female applicants to NTU's male-dominated STEM programs being disadvantaged by this type of tests, girls may be placed in a lower math track in middle or high school than their true learning ability would warrant because they underperform in the highly uncertain placement test. As the math track serves as a signal of ability to students, the likely female underperformance in placement

tests can lead girls to systematically underestimate their STEM aptitude and thus develop less interest in advanced STEM education and careers.

Because our results derive from application records from a single university in Taiwan, we must acknowledge a possible limitation in the study's generalizability. Nevertheless, other East Asian countries also have a handful of incontestably top universities such as NTU as well as comparable admission systems (Marginson 2011; Wang et al. 2024; Seoul National University 2025). Therefore, we think that this study does offer general lessons about the obstacles the most high-achieving STEM-oriented women and girls may encounter in other countries. Besides, as mentioned earlier, our use of NTU instead of general college application records for the analysis is actually beneficial, because it enables us to better account for gender disparities in STEM aptitude and exposure, which are likely much greater among the general college applicants than among the very ambitious and gifted ones who aspire to enter male-concentrated elite STEM programs at NTU.

We should also caution that institutional gatekeepers and their evaluation criteria only explain one piece of the puzzle on women's underrepresentation in elite STEM programs, as NTU's male-concentrated programs have far fewer female than male applications to begin with. Our results about assessment methods nevertheless have broader implications beyond gender and STEM education. Prior research has long called attention to the female disadvantage in standardized tests (Niederle and Vesterlund 2010), but we have shown that the type of tests with even less well-defined boundaries deserve equal, if not more, attention. At the same time, we have found female advantages in written application materials (e.g., essays) and interviews, although the extent of the advantage varies by the area of study. Given that male and female students perform differently according to the assessment type, future researchers should focus more on how well results from each assessment type predict future school and career success. Instead of adopting certain evaluation methods out of inertia or the attempt to imitate peers, gatekeepers need to be more circumspect about any method's validity to avoid unnecessarily enlarging gender inequality in educational and occupational opportunities.

Notes

- ¹ The number of seats for each department is determined by the Ministry of Education in Taiwan based on the state's long-term projection of skill demands and tends to vary little over time at older elite colleges such as NTU.
- ² With the GSAT administered just once a year, students can easily calculate their first-round chances based on their GSAT percentile and would not bother preparing a dossier for an NTU program if they fall below a certain percentile. This is also why out of a total of 350,000 Taiwanese students who graduated from high school in 2022 and 2023, only 4.6 percent applied to any NTU departments, even though 95 percent of them advanced to college (Ministry of Education, R.O.C. various years).
- ³ Other than those eliminated at the first round, 2,276 eligible applications were withdrawn upon passing the first round and excluded from our analytic sample as a result. Because people can apply to up to five NTU departments and be accepted into the second round by multiple programs simultaneously, some may give up a second-round opportunity

to focus their energy on the departments they have better chances of being admitted. We also excluded 385 applications because they were submitted to an interdisciplinary program for which few departmental characteristics are available; they were submitted through a special channel that reserved quota for disadvantaged students; or they missed key information such as high school name or home address.

- 4 The applications with missing GPA from our data still have that information factored in during the application process because evaluators of the application dossiers would include a physical copy of the transcript; it is just the transcript record was not imported into our data set.
- 5 We ranked applications based the department's advertised GSAT preferences for the first round. Departments could also weigh subjects of the national entrance exam differently to determine admissions prior to admission reforms.
- 6 Because departments may weigh different GSAT subject scores differently, we calculated the maximum GSAT score an applicant can have and use the applicant's actual weighted total score divided by the maximum possible score to standardize the GSAT results across departments.
- 7 Although the average lengths of previous academic works may vary by the area of interest, our use of department fixed effects accounts for any disciplinary differences in the type and length of writing samples submitted.
- 8 However, a further analysis indicated no gender gap in the GSAT scores among applications within each department.
- 9 The results are similar, but the gender differences are more pronounced, when we do not include high school GPA or dossier length in the models (see Table S6 in the online supplement).

References

- Alon, Sigal and Marta Tienda. 2007. "Diversity, Opportunity, and the Shifting Meritocracy in Higher Education." *American Sociological Review* 72(4):487–511. <https://doi.org/10.1177/000312240707200401>
- Bahar, A. Kadir. 2021. "Will We Ever Close the Gender Gap among Top Mathematics Achievers? Analysis of Recent Trends by Race in Advanced Placement (Ap) Exams." *Journal for the Education of the Gifted* 44(4):331–65. <https://doi.org/10.1177/01623532211044540>
- Bastedo, Michael. 2021. "Holistic Admissions as a Global Phenomenon." Pp. 91–114 in *Higher Education in the Next Decade*. Brill. https://doi.org/10.1163/9789004462717_006
- Belasco, Andrew S., Kelly O. Rosinger, and James C. Hearn. 2015. "The Test-Optional Movement at America's Selective Liberal Arts Colleges: A Boon for Equity or Something Else?" *Educational Evaluation and Policy Analysis* 37(2):206–23. <https://doi.org/10.3102/0162373714537350>
- 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. <https://doi.org/10.3102/00028312211003526>
- Bielby, Rob, Julie Renee Posselt, Ozan Jaquette, and Michael N. Bastedo. 2014. "Why Are Women Underrepresented in Elite Colleges and Universities? A Non-Linear Decomposition Analysis." *Research in Higher Education* 55(8):735–60. <https://doi.org/10.1007/s11162-014-9334-y>

- Bordon, Paola and Chao Fu. 2015. "College-Major Choice to College-Then-Major Choice." *The Review of Economic Studies* 82(4):1247–88. <https://doi.org/10.1093/restud/rdv023>
- Brandell, Gerd and Else-Marie Staberg. 2008. "Mathematics: A Female, Male or Gender-Neutral Domain? A Study of Attitudes among Students at Secondary Level." *Gender and Education* 20(5):495–509. <https://doi.org/10.1080/09540250701805771>
- Brenøe, Anne Ardila and Ulf Zölitz. 2020. "Exposure to More Female Peers Widens the Gender Gap in Stem Participation." *Journal of Labor Economics* 38(4):1009–54. <https://doi.org/10.1086/706646>
- Brotman, Jennie S. and Felicia M. Moore. 2008. "Girls and Science: A Review of Four Themes in the Science Education Literature." *Journal of Research in Science Teaching* 45(9):971–1002. <https://doi.org/10.1002/tea.20241>
- Buchmann, Claudia, Thomas A. DiPrete, and Anne McDaniel. 2008. "Gender Inequalities in Education." *Annual Review of Sociology* 34(1):319–37. <https://doi.org/10.1146/annurev.soc.34.040507.134719>
- Charles, Maria and Karen Bradley. 2009. "Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries." *American Journal of Sociology* 114(4):924–76. <https://doi.org/10.1086/595942>
- Chen, Yi-Ling. 2020. "'Housing Prices Never Fall': The Development of Housing Finance in Taiwan." *Housing Policy Debate* 30(4):623–39. <https://doi.org/10.1080/10511482.2020.1714691>
- Coffman, Katherine B. and David Klinowski. 2020. "The Impact of Penalties for Wrong Answers on the Gender Gap in Test Scores." *Proceedings of the National Academy of Sciences* 117(16):8794–803. <https://doi.org/10.1073/pnas.1920945117>
- Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106(6):1691–730. <https://doi.org/10.1086/321299>
- Correll, Shelley J. 2004. "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations." *American Sociological Review* 69(1):93–113. <https://doi.org/10.1177/000312240406900106>
- Correll, Shelley J. and Cecilia L. Ridgeway. 2006. "Expectation States Theory." Pp. 29–51 in *Handbook of Social Psychology*, edited by J. Delamater. Boston, MA: Springer US. https://doi.org/10.1007/0-387-36921-X_2
- Devine, Amy, Kayleigh Fawcett, Dénes Szűcs, and Ann Dowker. 2012. "Gender Differences in Mathematics Anxiety and the Relation to Mathematics Performance While Controlling for Test Anxiety." *Behavioral and Brain Functions* 8(1):33. <https://doi.org/10.1186/1744-9081-8-33>
- Downey, Douglas B. and Anastasia S. Vogt Yuan. 2005. "Sex Differences in School Performance during High School: Puzzling Patterns and Possible Explanations." *The Sociological Quarterly* 46(2):299–321. <https://doi.org/10.1111/j.1533-8525.2005.00014.x>
- Felson, Richard B. and Lisa Trudeau. 1991. "Gender Differences in Mathematics Performance." *Social Psychology Quarterly* 54(2):113–26. <https://doi.org/10.2307/2786930>
- Gerber, Theodore P. and Sin Yi Cheung. 2008. "Horizontal Stratification in Postsecondary Education: Forms, Explanations, and Implications." *Annual Review of Sociology* 34(1):299–318. <https://doi.org/10.1146/annurev.soc.34.040507.134604>
- Gërxhani, Klarita, Jordi Brandts, and Arthur Schram. 2023. "Competition and Gender Inequality: A Comprehensive Analysis of Effects and Mechanisms." *American Journal of Sociology* 129(3):715–52. <https://doi.org/10.1086/727822>

- Global Views Magazine. 2024. "2024 Taiwan's Best University Ranking," Taipei: Global Views Magazine. Retrieved November 19, 2024 (<https://www.gvm.com.tw/article/113992>).
- Goldin, Claudia and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians." *The American Economic Review* 90(4):715–41. <https://doi.org/10.1257/aer.90.4.715>
- Graetz, Georg and Arizo Karimi. 2022. "Gender Gap Variation across Assessment Types: Explanations and Implications." *Economics of Education Review* 91:102313. <https://doi.org/10.1016/j.econedurev.2022.102313>
- Griffith, Amanda L. 2010. "Persistence of Women and Minorities in Stem Field Majors: Is It the School That Matters?" *Economics of Education Review* 29(6):911–22. <https://doi.org/10.1016/j.econedurev.2010.06.010>
- Grodsky, Eric, John Robert Warren, and Erika Felts. 2008. "Testing and Social Stratification in American Education." *Annual Review of Sociology* 34(1):385–404. <https://doi.org/10.1146/annurev.soc.34.040507.134711>
- Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. "Culture, Gender, and Math." *Science* 320(5880):1164–65. <https://doi.org/10.1126/science.1154094>
- Hamidi, Zachariah, Steven J Durning, Dario Torre, Robert Liotta, and Ting Dong. 2021. "Do Interviews Influence Admission Decisions? An Empirical Analysis from an Institution." *Military Medicine* 186(3–4):426–36. <https://doi.org/10.1093/milmed/usaa477>
- Hamilton, Laura S. 1998. "Gender Differences on High School Science Achievement Tests: Do Format and Content Matter?" *Educational Evaluation and Policy Analysis* 20(3):179–95. <https://doi.org/10.3102/01623737020003179>
- Hannum, Emily, Hiroshi Ishida, Hyunjoon Park, and Tony Tam. 2019. "Education in East Asian Societies: Postwar Expansion and the Evolution of Inequality." *Annual Review of Sociology* 45:625–47. <https://doi.org/10.1146/annurev-soc-073018-022507>
- Haveman, Heather A. and Lauren S. Beresford. 2012. "If You're So Smart, Why Aren't You the Boss? Explaining the Persistent Vertical Gender Gap in Management." *The Annals of the American Academy of Political and Social Science* 639:114–30. <https://doi.org/10.1177/0002716211418443>
- Hyde, Janet Shibley and Marcia C. Linn. 2006. "Gender Similarities in Mathematics and Science." *Science* 314(5799):599–600. <https://doi.org/10.1126/science.1132154>
- Hyde, Janet Shibley. 2014. "Gender Similarities and Differences." *Annual Review of Psychology* 65(1):373–98. <https://doi.org/10.1146/annurev-psych-010213-115057>
- Kahn, Shulamit and Donna Ginther. 2018. "Women and Science, Technology, Engineering, and Mathematics (Stem): Are Differences in Education and Careers Due to Stereotypes, Interests, or Family?" Pp. 0 in *The Oxford Handbook of Women and the Economy*, edited by S. L. Averett, L. M. Argys, and S. D. Hoffman. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190628963.013.13>
- Karabel, Jerome. 2005. *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton*. Houghton Mifflin Harcourt.
- Kariya, Takehiko. 2020. "Japan's Post-Catch-up Modernity: Educational Transformation and Its Unintended Consequences." Pp. 304–17 in *Routledge Handbook of Contemporary Japan*. Routledge. <https://doi.org/10.4324/9781315544748-22>
- Kimball, Meredith M. 1989. "A New Perspective on Women's Math Achievement." *Psychological Bulletin* 105(2):198–214. <https://doi.org/10.1037/0033-2909.105.2.198>.
- Leeper, Campbell and Carly Kay Friedman. 2007. "The Socialization of Gender." Pp. 561–587 in *Handbook of Socialization: Theory and Research*, edited by J. E. Grusec and P. D. Hastings. The Guilford Press.

- Lee, Jaekyung and Daekwon Park. 2014. "Do American and Korean Education Systems Converge? Tracking School Reform Policies and Outcomes in Korea and the USA." *Asia Pacific Education Review* 15(3):391–99. <https://doi.org/10.1007/s12564-014-9325-x>
- Legewie, Joscha and Thomas A. DiPrete. 2014. "The High School Environment and the Gender Gap in Science and Engineering." *Sociology of Education* 87(4):259–80. <https://doi.org/10.1177/0038040714547770>
- Li, Qing. 1999. "Teachers' Beliefs and Gender Differences in Mathematics: A Review." *Educational Research* 41(1):63–76. <https://doi.org/10.1080/0013188990410106>
- Li, Wangyang and Yu Xie. 2020. "The Influence of Family Background on Educational Expectations: A Comparative Study." *Chinese Sociological Review* 52(3):269–94. <https://doi.org/10.1080/21620555.2020.1738917>
- Liu, Ruo-Fan. 2022. "Pathways to College Admissions: Student Strategies and Class Variations in Activating Cultural Knowledge in Taiwan." *International Studies in Sociology of Education* 31(3):284–304. <https://doi.org/10.1080/09620214.2020.1852882>
- Luoh, Ming-Ching. 2002. "Who Are NTU Students?-Differences across Ethnic and Gender Groups and Urban/Rural Discrepancy." *Taiwan Economic Review* 30(1):113–46
- Ma, Liping, Xin Li, Qiong Zhu, and Xiaoyang Ye. 2023. "College-Major Choice to College-Then-Major Choice: Experimental Evidence from Chinese College Admissions Reforms." *Economics of Education Review* 94:102380. <https://doi.org/10.1016/j.econedurev.2023.102380>
- Machado, Cecilia and Christiane Szerman. 2021. "Centralized College Admissions and Student Composition." *Economics of Education Review* 85:102184. <https://doi.org/10.1016/j.econedurev.2021.102184>
- Marginson, Simon. 2011. "Higher Education in East Asia and Singapore: Rise of the Confucian Model." *Higher Education* 61(5):587–611. <https://doi.org/10.1007/s10734-010-9384-9>
- Marks, Gary N. 2008. "Accounting for the Gender Gaps in Student Performance in Reading and Mathematics: Evidence from 31 Countries." *Oxford Review of Education* 34(1):89–109. <https://doi.org/10.1080/03054980701565279>
- Ministry of Education, Republic of China (Taiwan). various years. "Basic Statistics of Schools". <https://depart.moe.edu.tw/ed4500/News.aspx?n=5A930C32CC6C3818&sms=91B3AAE8C6388B96>.
- Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." *Proceedings of the National Academy of Sciences* 109(41):16474–79. <https://doi.org/10.1073/pnas.1211286109>
- Niederle, Muriel and Lise Vesterlund. 2010. "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives* 24(2):129–44. <https://doi.org/10.1257/jep.24.2.129>
- O'Dea, R. E., M. Lagisz, M. D. Jennions, and S. Nakagawa. 2018. "Gender Differences in Individual Variation in Academic Grades Fail to Fit Expected Patterns for Stem." *Nature Communications* 9(1):3777. <https://doi.org/10.1038/s41467-018-06292-0>
- Petersen, Jennifer. 2018. "Gender Difference in Verbal Performance: A Meta-Analysis of United States State Performance Assessments." *Educational Psychology Review* 30(4):1269–81. <https://doi.org/10.1007/s10648-018-9450-x>
- Posselt, Julie, Kamaria B. Porter, and Aurora Kamimura. 2018. "Organizational Pathways toward Gender Equity in Doctoral Education: Chemistry and Civil Engineering Compared." *American Journal of Education* 124(4):383–410. <https://doi.org/10.1086/698457>

- Posselt, Julie R. 2015. "Disciplinary Logics in Doctoral Admissions: Understanding Patterns of Faculty Evaluation." *The Journal of Higher Education* 86(6):807–33. <https://doi.org/10.1080/00221546.2015.11777385>
- Quadlin, Natasha. 2020. "From Major Preferences to Major Choices: Gender and Logics of Major Choice." *Sociology of Education* 93(2):91–109. <https://doi.org/10.1177/0038040719887971>
- Raabe, Isabel J., Zsófia Boda, and Christoph Stadtfeld. 2019. "The Social Pipeline: How Friend Influence and Peer Exposure Widen the Stem Gender Gap." *Sociology of Education* 92(2):105–23. <https://doi.org/10.1177/0038040718824095>
- Ramirez, Francisco O. and Naejin Kwak. 2015. "Women's Enrollments in Stem in Higher Education: Cross-National Trends, 1970–2010." Pp. 9–49 in *Advancing Women in Science: An International Perspective*, edited by J. W. Pearson, L. M. Frehill, and C. L. McNeely. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-08629-3_2
- Ridgeway, Cecilia L. 2013. "Why Status Matters for Inequality." *American Sociological Review* 79(1):1–16. <https://doi.org/10.1177/0003122413515997>
- Ridgeway, Cecilia L. and Shelley J. Correll. 2004. "Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations." *Gender & Society* 18(4):510–31. <https://doi.org/10.1177/0891243204265269>
- Riegle-Crumb, Catherine and Menglu Peng. 2021. "Examining High School Students' Gendered Beliefs About Math: Predictors and Implications for Choice of Stem College Majors." *Sociology of Education* 94(3):227–48. <https://doi.org/10.1177/00380407211014777>
- Saboe, Matt and Sabrina Terrizzi. 2019. "Sat Optional Policies: Do They Influence Graduate Quality, Selectivity or Diversity?" *Economics Letters* 174:13–17. <https://doi.org/10.1016/j.econlet.2018.10.017>
- Sassler, Sharon, Jennifer Glass, Yael Levitte, and Katherine M. Michelmore. 2017. "The Missing Women in Stem? Assessing Gender Differentials in the Factors Associated with Transition to First Jobs." *Social Science Research* 63:192–208. <https://doi.org/10.1016/j.ssresearch.2016.09.014>
- Seoul National University. 2025, "Seoul National University Application Guide." Retrieved November 19, 2024 (<https://admission.snu.ac.kr/undergraduate/early/guide>).
- Stearns, Elizabeth, Martha Cecilia Bottía, Eleonora Davalos, Roslyn Arlin Mickelson, Stephanie Moller, and Lauren Valentino. 2016. "Demographic Characteristics of High School Math and Science Teachers and Girls' Success in Stem." *Social Problems* 63(1):87–110. <https://doi.org/10.1093/socpro/spv027>
- Stern, Robert M., Mary W. Montgomery, Nora Y. Osman, Joel T. Katz, and Maria A. Yialamas. 2021. "The Relationship between Applicant Gender and Internal Medicine Residency Interview Scores." *Journal of Graduate Medical Education* 13(6):814–21. <https://doi.org/10.4300/JGME-D-21-00270.1>
- Sullivan, Alice, Samantha Parsons, Francis Green, Richard D. Wiggins, and George Ploubidis. 2018. "Elite Universities, Fields of Study and Top Salaries: Which Degree Will Make You Rich?" *British Educational Research Journal* 44(4):663–80. <https://doi.org/10.1002/berj.3453>
- Times Higher Education, 2024, "World University Rankings." Retrieved November, 19, 2024 (<https://www.gvm.com.tw/article/113992>).
- Uhlmann, Eric Luis and Geoffrey L. Cohen. 2005. "Constructed Criteria: Redefining Merit to Justify Discrimination." *Psychological Science* 16(6):474–80. <https://doi.org/10.1111/j.0956-7976.2005.01559.x>

- Wang, Shutao, Lujing Tang, and Junwei Bao. 2024. "Comparison of Comprehensive Evaluation's and Unified Examination's Effectiveness in University Enrolment: Chinese Evidence from Propensity Score Matching Analysis." *Journal of Higher Education Policy and Management* 46(6):652–70. <https://doi.org/10.1080/1360080X.2024.2354535>
- Weeden, Kim A., Dafna Gelbgiser, and Stephen L. Morgan. 2020. "Pipeline Dreams: Occupational Plans and Gender Differences in Stem Major Persistence and Completion." *Sociology of Education* 93(4):297–314. <https://doi.org/10.1177/0038040720928484>
- Xie, Yu, Michael Fang, and Kimberlee Shauman. 2015. "Stem Education." *Annual Review of Sociology* 41:331–57. <https://doi.org/10.1146/annurev-soc-071312-145659>
- Yu, Wei-hsin. 2009. *Gendered Trajectories: Women, Work, and Social Change in Japan and Taiwan*. Stanford, CA: Stanford University Press. <https://doi.org/10.11126/stanford/9780804760096.001.0001>
- Yu, Wei-hsin and Kuo-Hsien Su. 2024. "Who Benefits from Elite Colleges' Decreased Reliance on High-Stakes Standardized Tests? Evidence from a Quasi-Field Experiment." *Research in Social Stratification and Mobility* 89:100871. <https://doi.org/10.1016/j.rssm.2023.100871>
- Zimmerman, Seth D. 2019. "Elite Colleges and Upward Mobility to Top Jobs and Top Incomes." *American Economic Review* 109(1):1–47. <https://doi.org/10.1257/aer.20171019>

Acknowledgments: We thank National Taiwan University for sharing application records with the authors for the purpose of academic research. We also acknowledge the valuable input from Yu Xie at an earlier stage of this research project and a grant from the Asia Pacific Center at UCLA awarded to the first author.

Wei-hsin Yu: Department of Sociology, University of California, Los Angeles.
E-mail: whyu@soc.ucla.edu.

Kuo-Hsien Su: Department of Sociology, National Taiwan University.
E-mail: khsu@ntu.edu.tw.