Missing stars?

Quantifying the gender gap in the assessment of gifted students*

Beatrix Eugster†, Kelli Marquardt‡, Aurélien Sallin§

February 2025

PRELIMINARY DRAFT- Please do not cite or circulate without authors' permission.

Abstract

Female students are less likely to be identified as intellectually gifted than male students when the identification process follows a two-stage procedure, i.e., where nomination for an assessment by teachers or parents precedes the expert assessment. We show that two mechanisms drive the gender gap. First, at any level of IQ - but especially at high levels of IQ -, male students are more likely to be nominated for assessment due to higher levels of externalizing behaviors. This difference accounts for roughly 70% of the gender gap. Second, but less importantly, the referral threshold in terms of externalizing behavior is lower for male students than for female students. We discuss potential policy implications.

Keywords: gifted education, two-stage diagnosis, gender gap

JEL Classification: I21, I24, J16

*We acknowledge data provision from the Ministry of Education and Nicole Wellinger of the Canton of St. Gallen. The views expressed here do not represent those of any current or past employer of the authors. Errors are our own.

†University of St.Gallen, CESifo, CEPR
‡Federal Reserve Bank of Chicago
§University of St.Gallen, SWICA. Corresponding author: aurelien.sallin@unisg.ch

1 Introduction

Extensive research documents significant gender gaps in employment, earnings, and promotions. Numerous mechanisms are proposed to account for the persistent gender gap, ranging from gender preferences for job characteristics (Caldwell and Danieli, 2024; Wasserman, 2023; Wiswall and Zafar, 2018) or gender identity norms (Bertrand, 2011; Blau and Kahn, 2017) to intrafamily labor markets (Bertrand, Kamenica, and Pan, 2015; Calvo, Lindenlaub, and Reynoso, 2024; Cortés and Pan, 2023). An important strand of literature argues that the mechanisms of gender differences in the economy begin to develop in early childhood, especially within education systems and for academic-related tasks (e.g., Exley and Kessler, 2022).

Despite progress in reducing gender gaps in achievement and graduation rates, there is a persistent gender gap in identifying gifted children, particularly disadvantaging gifted female students (Balestra, Sallin, and Wolter, 2023; Petersen, 2013). The issue of underidentifying intellectually gifted female students is concerning as it limits their access to adapted education programs and benefits, potentially widening long-term gender disparities in the economy. The underidentification of female students (or minority students more generally) in gifted programs occurs for two reasons. First, universal screening is not applied, meaning that only a subset of children will be assessed for placement in gifted programs. Card and Giuliano (2016b) show for instance that the absence of universal screening leads to under-representation of economically disadvantaged students in gifted programs. Second, children who are assessed for giftedness and placed in gifted education programs are endogenously selected, often following referral processes that exacerbate the disparity in giftedness identification. The standard process relies on teacher referrals, which, due to time and attention constraints in the classroom, may lead to referral decisions based on easily observable behaviors that are only partially correlated with children's underlying IQ (e.g., classroom disruption).

In this paper, we argue that two-stage referral processes can exacerbate the gender gap in giftedness identification due to either gender differences in (implicit) standards set by teachers to be referred for diagnosis and/or gender differences in the underlying correlation between classroom disruption and IQ. In two-step identification processes, children are first referred to an assessor from parents and teachers through a nomination stage. In the second step, an assessment is given and the child is either identified as intellectually gifted or not (see, for example Card and Giuliano, 2016b). We present a simplified model of this process to demonstrate

that the underidentification of female gifted students stems from two potential mechanisms: (1) gender differences in the underlying relationship between giftedness and classroom behavior/disruption, and (2) variations in teacher's tolerance for classroom disruption by gender. We base our model on the assumption that the IQ distribution is the same for female and male students, and we show that our conclusions hold even when we allow for gender-specific IQ distributions.

Using rich data on all children referred for an assessment to the school psychologists over 4 cohorts from a single educational system in Switzerland, we quantify the gender gap in giftedness identification. We then use these data to estimate parameters of a simple model that allows us to decompose the two mechanisms that we predict contribute to the gender gap in giftedness identification[1] We identify our parameter values using variation in referral rates across schools, within schools across gender, and the relationship between referral rates and average disruption measures conditional on referral.

Our findings suggest that gender differences in classroom behaviors are the main driver of the gender gap in giftedness identification. We show that at any level of IQ, male students are more likely to be referred for giftedness assessment due to higher rates of classroom behaviors that might trigger attention and response from teachers (in line with Bertrand and Pan, 2013). Specifically, at the threshold of intellectual giftedness (IQ of 130), males are over 2.5 times more likely than females to be referred and subsequently identified as gifted. In our decomposition exercise, we show that less than one-fifth of this gap is driven by gender differences in the thresholds that teachers use for student referral. Instead, the majority of the gap (about 70%) is explained by underlying gender differences in external behaviors (which we call classroom "disruption", see Lazear, 2001), and in the correlation between these behaviors and individual IQ. These results suggest a need to revisit the referral and assessment process, advocating for universal screening or alternative metrics for referral to ensure more equitable identification and support for gifted students.

This paper contributes to the research on gifted students in schools and provides an explanation for the widely documented fact that female students are less likely to be identified as gifted (Balestra, Sallin, and Wolter, 2023; Petersen, 2013), and less likely to be referred to a gifted program (Bianco et al., 2011)[2]. This is concerning, as most education systems within the

[1]The model and estimation strategies are in spirit of Marquardt (2023), which quantifies the mechanisms leading to gender differences in diagnosis rates for Attention Deficit Hyperactivity Disorder (ADHD).

[2]It has been argued that teacher referral is more prone to favor male students over female students in giftedness

OECD and in the US consider gifted students as a group with distinct educational needs and as a valuable source of human capital (Rutigliano and Quarshie, 2021a). Consequently, gifted students receive customized educational programs to meet their specific needs (Booij, Haan, and Plug, 2016; Bui, Craig, and Imberman, 2014; Card, Chyn, and Giuliano, 2024; Card and Giuliano, 2016a; Lavy and Goldstein, 2022).[3] We show that female students are likely to miss out on these customized education programs / benefits and document the mechanisms that contribute to the disparity in early education.

Our paper refines our understanding of the limitations inherent in two-stage assessment processes commonly used in education and behavioral health. This process has been shown to systematically convey gender, race, or socioeconomic disparities, and thus miss many potentially qualified students (Card and Giuliano, 2016b; Cohen, 2022; Harris, Hill, and Lenard, 2022; McBee, Peters, and Miller, 2016).[4] This phenomenon is also evident in special education identification and ADHD diagnosis. For example, Elder et al. (2021) highlight racial and ethnic disparities in the identification of special education, attributing the disparities to the salience of behaviors and subjective comparisons with the level of racial composition at the referral stage. In the diagnosis of ADHD, research indicates that the birth date of a child relative to the school entry cutoff significantly influences the diagnosis rates, suggesting that teachers make referral decisions based on a subjective comparison of the behavior of young students with that of older and more mature students (Balestra, Eugster, and Liebert, 2020; Elder, 2010; Layton et al., 2018; Persson, Qiu, and Rossin-Slater, 2021; Schwandt and Wuppermann, 2016). Similarly, Marquardt (2023) studies the gender gap in ADHD diagnosis using a similar two-stage model as we describe in this paper. She finds that subjective assessment in the second stage is a primary driver of the gender gap in the identification of ADHD. Our paper contributes to this literature by illustrating how classroom behaviors influence the two-stage assessment process for giftedness identification.

Finally, this paper adds to the literature on the gender gap in the economy more broadly, as cited above. Although there has been significant progress in reducing gender gaps in achievement

identification, while (especially at lower grade) grade point averages are more likely to flag female students as gifted (Petersen, 2013).

[3]Although gifted education varies significantly from US state to US state, and even within districts in a single state, and although there is no federal mandate related to gifted identification and services, Gifted and Talented (GT) education programs in the US serve more than 3 million public school students (Harris, Hill, and Lenard, 2022).

[4]For a comprehensive discussion on the two-stage identification process for gifted students in the U.S., see McBee, Peters, and Miller (2016).

and graduation rates, e.g., through role models and the reduction of teacher gender stereotypes (Carlana, 2019; Carrell, Page, and West, 2010), gifted female students remain underidentified as gifted students and thus miss out on important resources which can have long-term implications on their economic and social outcomes. Thus, our paper provides additional insight into why a significant gender gap in employment, earnings, and promotions persists.

The remainder of the paper is structured as follows. We start by describing background details for giftedness identification and how the process works in our application setting as well as the available data in Section 2. In Section 3 we present a simple theoretical framework that highlights the mechanisms that may contribute to disparities and underidentification of giftedness in practice. We outline the empirical model and estimation strategy in Section 4. Section 5 summarizes our results with a simulated decomposition exercise, and section 6 concludes.

2 Background, Setting, and Data

Background: giftedness assessment over the world Intellectually gifted children or students are defined as "children, students, or youth who give evidence of high-performance capability in areas such as intellectual, creative, artistic, or leadership capacity, or in specific academic fields, and who require services or activities not ordinarily provided by the school in order to fully develop such capabilities."[5] The concept of giftedness is traditionally associated with high intelligence and exceptional performance. It is measured with IQ testing and IQ scores, which are normally distributed with a mean of 100 and a standard deviation of 15. An IQ score cutscore of 130 and above (two standard deviations above the mean) (McClain and Pfeiffer, 2012) is used to diagnose individuals as gifted. This means that around 2.5% of the population is gifted as defined by IQ. In recent decades, however, many US states and European countries have included other dimensions in their definition of giftedness beyond advanced intellectual ability, such as academic performance, creativity and creative thinking, arts and leadership (Rinn, Mun, and Hodges, 2022). The US states have the authority to determine how to define, identify, and serve gifted students (Hafenstein, Boley, and Lin, 2022; McClain and Pfeiffer, 2012; Rutigliano and Quarshie, 2021b; VanTassel-Baska, 2018). In 2021,

[5]US Federal government statutory definition of gifted students in the Improving America's Schools Act of 1994, P.L. 103-382, Title XIV, p. 388, https://web.archive.org/web/20050308112728/http://www.ed.gov/legislation/ESEA/sec14101.html

46 US states have a definition of "gifted" in law or policy, whereas 41 require by law or rule the identification of gifted and talented students (Rinn, Mun, and Hodges, 2022).

The landscape of gifted education in OECD countries is characterized by a lack of uniform policy and the underrepresentation of minorities in gifted programs. In the U.S., gifted education varies significantly from state to state, and even within districts in a single state. There is no federal mandate related to gifted identification and services. Therefore, states have the authority to determine how to define, identify, and serve gifted students (Hafenstein, Boley, and Lin, 2022; McClain and Pfeiffer, 2012; Rutigliano and Quarshie, 2021b; VanTassel-Baska, 2018).

Standard processes for identifying gifted students often rely on a two-stage identification process. For instance, US states have various ways of identifying gifted students. In 2011, 16 (17) states mandate that schools use intelligence tests (achievement tests) to identify gifted students, while 13 states require the use of nominations (teacher and/or parent referrals) and 9 states require the use of a teacher-completed behavior rating scale (McClain and Pfeiffer, 2012). In Europe, the process varies across countries and often relies on nominations as well (Rutigliano and Quarshie, 2021b). In the first, nomination stage, children are referred to an assessor from parents and teachers. In the second step, a psychologist makes an assessment of the giftedness of the child (for a complete discussion of the two stage identification process for gifted students in the U.S., see McBee, Peters, and Miller, 2016).

This referral and identification process tends to miss many qualified students, especially from minority and economically disadvantaged groups (Card and Giuliano, 2016b; McBee, Peters, and Miller, 2016). Inequity in gifted assessment and access to gifted education has been a source of controversy, with arguments that it widens the gap between high-achieving students and their peers, particularly along racial lines (Cohen, 2022). An important disparity in giftedness assessment occurs between female and male students. Because referral is the first stage for the identification of giftedness, this contributes to the underidentification of female students: male students are more often referred for psychological assessment because they are on average more disruptive than female students. Following Lazear (2001), gender differences in the display of "disruptive" behaviors explain differences in the identification of giftedness. These behaviors cause the teacher to allocate her time to these students rather than to the other students in the classroom.[6] In this article, we provide evidence that male children have higher measures of

[6]Note that disruption does not have to be negative; students asking very complicated questions can be as disruptive as students misbehaving.

disruption on average, with a larger gap at the higher end of the IQ distribution. In addition, giftedness is more often correlated with disruptive behaviors (Brown, Reichel, and Quinlan, 2009; Guénolé et al., 2013; Tasca et al., 2022).

Setting: giftedness assessment in St. Gallen, Switzerland We build on the literature on two-stage identification processes for gifted students using data on all students enrolled in the primary and secondary schools of the Swiss canton of St. Gallen (around 500,000 inhabitants) (Balestra, Eugster, and Liebert, 2020, 2022; Balestra, Sallin, and Wolter, 2023). In this setting, students with special educational needs, including gifted students, are identified and assessed during two-stage identification processes. First, students are nominated for assessment by the teacher to the School Psychological Service (SPS), a centralized "gatekeeper" that assesses special educational needs and refers the students for further treatment. The SPS is responsible for identifying and providing psychological support to gifted children. For most students (about nine out of ten), the SPS services are requested directly by the teacher, but some requests are also submitted by parents or the child's doctor. The SPS is centralized at the regional level, and all regional offices serve the schools in their region. Students can be referred at any time during their mandatory schooling, with an average age at referral of almost 9 years. During the first visit to the psychologist, the students undergo an assessment under the supervision of a trained psychologist. During this assessment, their cognitive abilities are assessed with an IQ test.

The canton of St. Gallen bears the responsibility for the education of children identified as gifted and must guarantee the fulfillment of their educational needs. In this regard, emphasis is put on the inclusion of students identified as gifted in regular classrooms (mainstreaming). Requests to send gifted children to special schools are accepted only under strict conditions: the child must have already skipped a grade, justify why the classroom environment is not adequate, and undergo a psychological evaluation. In the vast majority of cases, special activities, additional support and enrichment programs are offered outside of class, depending on the school and upon request of parents and teachers. Acceleration (skipping a grade) or starting school at a younger age are also possible in some rare cases.

Data We combine data on schools and cohort sizes from the Ministry of Education of the Canton of St. Gallen with data on students referred to the services of the School Psychological Service of the Canton of St. Gallen (SPS). For further background information on the data and

Table 1: Summary statistics

	Average	Min.	Max.	Sd.	N
All students, school-cohort level ($N = 12,214$)					
Cohort size	117.44	64.00	503.00	86.77	
Rate referral to the SPS	0.30	0.10	0.54	0.09	
Rate referral to the SPS male	0.34	0.09	0.65	0.11	
Rate referral to the SPS female	0.26	0.03	0.51	0.09	
Students, conditional on referral to the SPS ($N = 4,002$)					
Female	0.44	0	1	0.50	
Age at referral	8.68	3.94	15.45	2.09	
Number of contacts	7.16	1	85.00	6.21	
Tested for IQ	0.72	0	1	0.45	
IQ	97.71	42	145	12.27	
Identified as gifted	0.04	0	1	0.20	164
Identified as gifted: female students	0.01	0	1	0.17	48
Identified as gifted: male students	0.03	0	1	0.17	116

Notes: This table reports summary statistics about our sample. All students born between 1997 and 2000 and referred to the School Psychological Services (SPS) in the Swiss Canton of St. Gallen are included in the sample. 104 distinct school-cohorts. *Source*: Ministry of Education, Canton of St. Gallen

the setting, see Balestra, Eugster, and Liebert (2020, 2022) and Eugster (2017).[7] Our resulting data include all students born between 1997 and 2000 referred to SPS in 104 different school cohorts (26 schools) of the Swiss Canton of St. Gallen. The average cohort size per school is 117 students.

Schools are organized at the municipality level, and children are assigned to schools on the sole basis of their location of residence. This strict assignment procedure is thoroughly implemented so that parents have no say about their child's school other than moving permanently to a different municipality or enrolling their children in a private school. Despite this rule, private schooling remains very rare in Switzerland. As the 2018 Education Report by the Swiss Coordination Center for Research in Education shows, more than 95 percent of children in 2016 attended public-funded schools in their municipality of residence.

The 4,002 students referred to the SPS represent around a third of the overall student population (N = 12,214) in these cohorts. Table 1 provides descriptive statistics. For all students referred to the SPS, we are able to observe their gender, age, school of referral, reason for referral, person who initiated the referral, and all the text written by the caseworker. The most common reasons for referral to the SPS are learning disabilities and emotional/social/behavioral

[7]The data preparation, empirical setting and variables are the same as used by Balestra, Sallin, and Wolter (2023) in their study on gifted students in Switzerland. We refer to their study for more details on the institutional setting. For this paper, we additionally removed 92 school-cohorts which either have a referral rate of zero ("never-referring") or one ("always-referring").

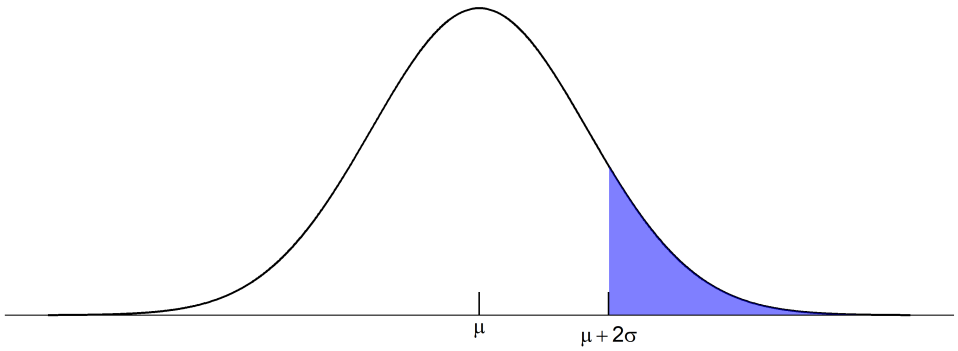
problems.

Most importantly, we observe a measure of IQ for most of the students referred to the SPS.[8] We classify a child as gifted if the IQ score is equal to or above 130, or if the psychologist qualitatively assessed the student to be gifted. This threshold (i.e. two standard deviations above the mean) is standard in the identification of gifted students. We use the text written by the psychologists to confirm that the professional identification corresponds to the IQ score obtained. As stated in Balestra, Sallin, and Wolter (2023), our data allow us to mitigate the potential limits of IQ as a single measure of giftedness, as we also observe qualitative assessments of cognitive ability obtained from the diagnoses and comments of the psychologist. Qualitative assessments reliably complement quantitative scores and account for other dimensions of intelligence not assessed by the IQ test.[9] In total, 164 students were identified as gifted, of which 116 were male and 48 were female.

3 Theoretical Framework and Motivation

In this section, we present a theoretical framework that, together with descriptive facts, motivates an empirical model and estimation strategy to decompose the mechanisms contributing to the gender gap in giftedness identification.

Figure 1



Theoretical framework To start, let v_i be an individual's IQ that follows an arbitrary distribution with mean μ and standard deviation σ. Let G_i be an indicator of whether the

[8] To measure the IQ score, we proceed as in Balestra, Sallin, and Wolter (2023) Some children were tested multiple times, at different ages or with different testing procedures. We select the highest IQ score recorded per child

[9] We have access to all the text written by the psychologist after every meeting with the students. This includes qualitative assessments of intellectual giftedness. We extract this information from text data following the procedure explained in Balestra, Sallin, and Wolter (2023).

individual is considered "intellectually gifted" according to the formal definition (i.e., those with an IQ score at least two standard deviations above the mean). Formally, $G_i = 1(v_i \geq \mu + 2\sigma)$. The share of the population with $G_i = 1$ corresponds to the shaded blue region in Figure 1.

Because everyone who is referred by a teacher is also given an IQ test and an assessment of giftedness, the probability that an individual is identified as gifted is equal to the probability that they are referred and gifted. Note that in this example, we assume that the IQ test is not the final assessment tool for giftedness identification. This assumption is not too unrealistic in practice, as psychologists also use qualitative assessments with IQ test to formally test for giftedness. Formally, let R_i be an indicator for whether the student is referred and let IG_i be an indicator for whether the individual is identified as gifted. Following Bayes' rule, the probability that a student is identified as gifted is written as follows: $P(IG_i = 1) = P(R_i = 1 \cap G_i = 1) = P(R_i = 1 \mid G_i = 1) \times P(G_i = 1)$.

Now, suppose that IQ is both observed *and* used as the referral mechanism by teachers so that a teacher refers a student whenever the student's IQ is above a referral threshold τ, corresponding to the striped regions in Figure 2. Then $R_i = 1(v_i > \tau)$, and so long as $\tau < \mu + 2\sigma$, we would have $P(IG_i) = 1 \times P(G_i = 1)$ so all gifted students will be appropriately identified as such (corresponding to subfigure 2a).[10] Note that this is true under universal screening where $\tau = -\infty$ but also for any referral threshold at or below $\mu + 2\sigma$. On the other hand, "Missing Stars" occur when $\tau > \mu + 2\sigma$ so $P(R_i = 1 \mid G_i = 1) < 1$ and thus $P(IG_i = 1) < 1 \times P(G_i = 1)$. The share of missing starts is denoted by the blue area without stripes in subfigure 2b.

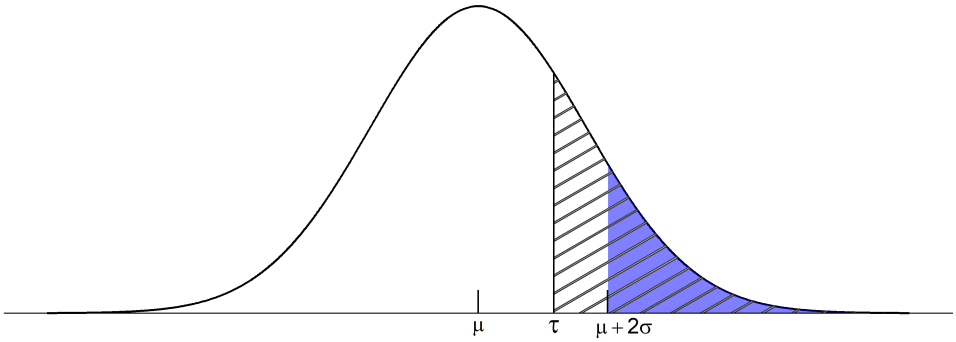
In this simple scenario, we see that the probability that an individual is identified as gifted is a function of both referral rules (striped area) and the underlying distribution of true IQ (blue shaded area). Mathematically, $P(IG_i = 1) = F(\tau, \mu, \sigma)$. Therefore, gender differences in giftedness identification can arise due to gender differences in referral thresholds, τ, and/or gender differences in the underlying IQ distribution parameters, μ and σ.

Assumption 1: males (m) and females (f) have identical IQ distributions so that $\mu_m = \mu_f$ and $\sigma_m = \sigma_f$.

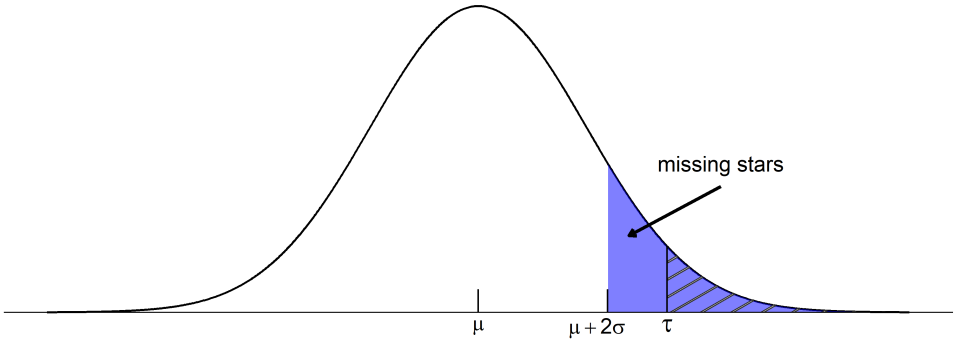
Under assumption 1, which we have argued is supported by the literature, the gender gap in gifted identification would be entirely driven by differences in τ such that $\tau_m \neq \tau_f$. In other words, in the situation where IQ was observed and used as the referral mechanism, then the

[10] This follows from Bayes Rule and the fact that whenever $\tau < \mu + 2\sigma$, any $v_i > \mu + 2\sigma$ implies $v_i > \tau$ so all gifted students will be referred with certainty.

Figure 2



(a)



(b)

gender gap in giftedness identification is explained by differences in the referral thresholds used by teachers. As detailed in the figures above, this will only lead to unwarranted disparities in giftedness identification in cases where at least one of the referral thresholds falls above $\mu + 2\sigma$.

However, teachers do not observe IQ before the assessment *nor* is IQ the referral mechanism. Instead, teachers must use other observable factors, such as salient disruptive behaviors, as decision rules. And these behaviors, called x_i, are probably a function of an individual's IQ such that $x_i = f(v_i)$.

In this more realistic scenario, a teacher will refer a student for psychological services if the observable factor, x_i, is greater than their tolerance threshold $\tilde{\tau}$. As before, $P(IG_i = 1) = P(R_i = 1 \mid G_i = 1) \times P(G_i = 1)$, but now $P(R_i = 1 \mid G_i = 1) = P(x_i > \tilde{\tau} \mid v_i > 2\sigma) = P(f(v_i) > \tilde{\tau} \mid v_i > 2\sigma)$. So, in addition to referral thresholds and IQ distribution parameters, giftedness identification will also depend on the relationship between IQ, v_i, and disruptive behaviors x_i, via the function $f(\cdot)$.

Again, under Assumption 1, we have that male and female students can be identified as gifted at different rates due to differences in the referral threshold (here $\tilde{\tau_m} \neq \tilde{\tau_f}$). However,

it is also plausible that there are gender differences in the IQ-disruption relationship such that $f_m(\cdot) \neq f_f(\cdot)$. In other words, even if a male and a female student have the same IQ score v_i, differences in the mapping of v_i to x_i by gender may further exacerbate differences in referral rates and giftedness identification above and beyond differences in teacher referral thresholds.

Formally, we argue that the gender gap in giftedness identification is driven by gender differences in referral thresholds ($\tilde{\tau_m} \neq \tilde{\tau_f}$) and/or gender differences in the IQ-disruption relationship, $f_m(\cdot) \neq f_f(\cdot)$. In the next paragraph, we present motivating facts from the data that help us link this conceptual framework to an empirical model for estimation.

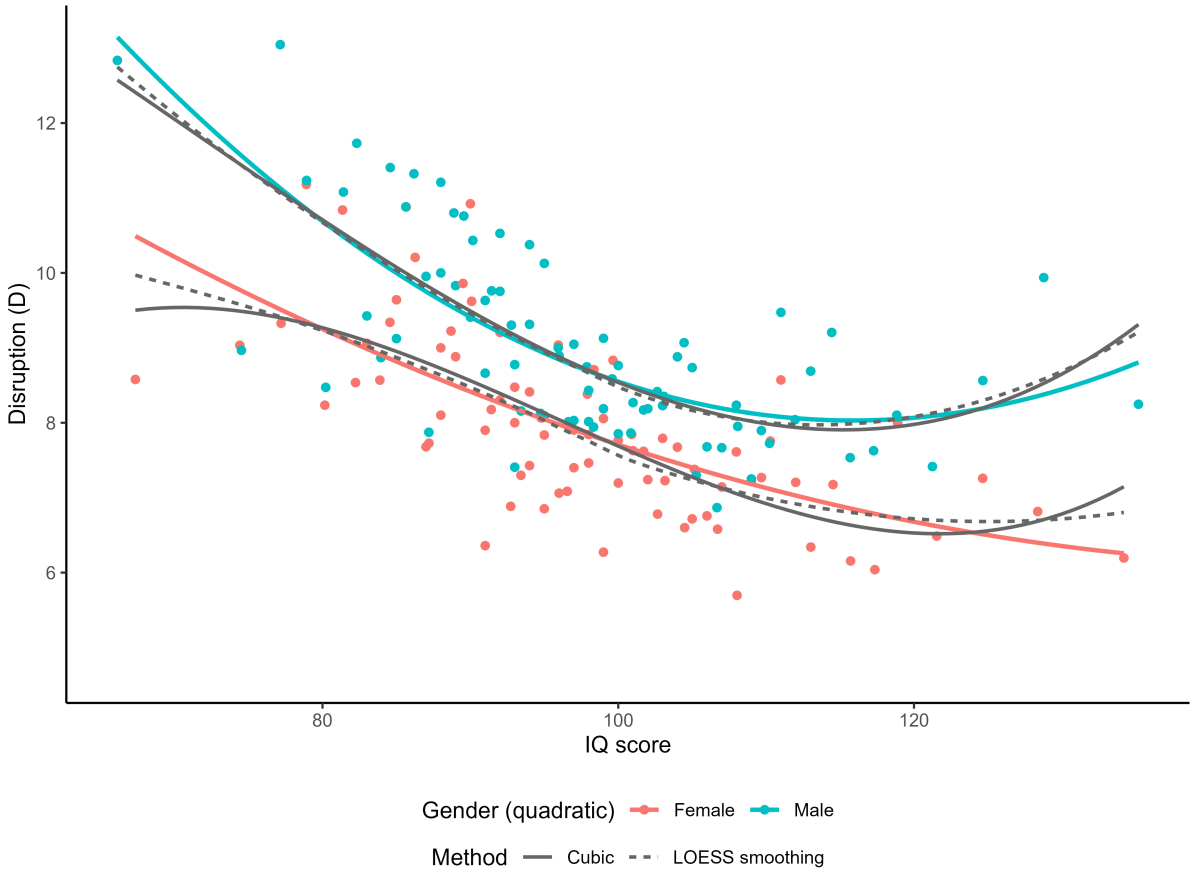
Motivating facts There are three facts about gender differences in giftedness identification that can be seen in our data and used to motivate our empirical approach. These were first observed in the data used by Balestra, Sallin, and Wolter (2023) in their study on the situation and effect of gifted peers in mandatory schools in Switzerland. First, we observe a gender gap in the number of identified gifted students. We observe around twice as many gifted male students as gifted female students (see Table 1), a ratio that diverges from expectations based on the general population. Given the overall distribution of IQ in the population, which we assume for now is not gender specific, this shows that many female students are missing and go through mandatory school without being identified as gifted. We posit that some of this difference is due to the difference in the referral rate to the school psychological service for female and male students: the referral rate for male students is 34%, while the referral rate for female students is 26%.

Second, we observe substantial variation in referral rates across schools. More importantly, we see that these referral rates are negatively correlated with the gender gap in giftedness identification. We argue that as the referral/tolerance threshold used by teachers decreases, more gifted students are referred. This in turn reduces the issue of underidentification and/or unwarranted disparities. We will use this correlation and variation in our empirical model.

Third, we show that, conditional on referral, the relationship between disruption and IQ follows a U-shape (Figure 3). Following Balestra, Eugster, and Liebert (2022), we define "disruption" as the severity of the case, which is proxied by the number of visits to the caseworker.[11] We observe that students in the lower distribution of IQ display more severe cases and that the

[11]Students who have been referred for problems related to emotional and social behaviors have around 50% more visits to the caseworker as students referred for other reasons, and students referred for learning disabilities have around 60% more visits.

Figure 3: IQ-Disruption Relationship (OLS) in the full sample



Note: This figure plots the gender specific IQ-disruption relationship for the set of students referred to psychological services for the full sample. Disruption is defined as the severity of the case, proxied by the number of visits to the caseworker. Observations are binned into 80 equally sized bins. Curves are determined via a ordinary least squares regression of disruption on standardized IQ and IQ squared, separately for male students and female students. We use a quadratic regression in our model, which is comparable to LOESS smoothing and cubic regression. The corresponding OLS coefficients and standard errors are presented in Appendix table X.

correlation between disruption and measured IQ is negative until a certain point. This correlation becomes positive after a given IQ threshold for male students, while it remains negative for gifted female students. Note that this relationship is robust to outliers in the number of visits. Gifted students have more severe cases than their peers in the middle of the IQ distribution. Male students have higher levels of disruption at any level of IQ.

4 Empirical Model and Estimation

In this section, we now turn to an empirical model that we can estimate and then simulate to quantify the mechanisms that contribute to the observed gender gap in giftedness identification. To do this, we add structure to the conceptual framework in Section 3 in a way that is consistent

with the motivating facts presented above. We account for both gender differences in the disruption-IQ relationship and variation in referral rates (selection into giftedness assessment). The setup and estimation of the model closely follow that in Marquardt (2023), which quantifies the mechanisms leading to gender differences in the diagnosis of ADHD.

Let z_i be our standardized IQ distribution following the distribution $N(0,1)$. Note that this rescaling means that students are intellectually gifted if they have a $z_i > 2$ (i.e. are two standard deviations above the mean). Following the literature, this underlying distribution is the same for male and female students. In Appendix B, we show that our main conclusions do not change when we allow for gender-specific IQ distributions (following the discussion on the greater male variability hypothesis Nyborg, 2015).

Let x_i be our measure of classroom disruption. For estimation purposes, we rescale the disruption measure shown in Figure 3 to be between 0 and 1.[12] Following the motivating figure 3, and observing the similarity between the semi-parametric and the parametric fits, we model the relationship between IQ and disruption, $f(\cdot)$, as quadratic. We also allow for the function parameters to differ by gender. Specifically, for $\theta \in \{m, f\}$ and standard normal error ε_i we model this relationship below. This standard error can be thought of as anything that impacts the individual's classroom behaviors in ways uncorrelated to their underlying IQ (e.g. family situations, cultural behaviors, etc.).

$$x_i|\theta_i = f_\theta(z_i) = \alpha_{0,\theta} + \alpha_{1,\theta} z_i + \alpha_{2,\theta} z_i^2 + \varepsilon_i \tag{1}$$

Recall, children are referred by their teachers to a school psychologist for an assessment if their disruption is above some threshold. We can think of the threshold as a "tolerance threshold", where teachers/schools have some level of behavior that they will "tolerate", and once these behaviors reach the limit, they will refer the student to the psychological services. In addition to allowing these tolerance thresholds to differ based on the gender of the student, we also allow them to vary across schools. This variation also helps us identify the $\alpha's$ in the IQ-disruption relationship since neither x_i nor z_i are observed by the econometrician unless a student is referred. Thus, schools with a low tolerance threshold will have higher referral rates, and marginally referred students will have disruption and IQ closer to the full population than those students from schools with a higher tolerance threshold.

[12] Specifically, we rescale our measure of disruption between 0 and 1 by first referral year.

Formally, the teacher observes x_i and will refer the child to a school psychologist for an assessment if this disruption is above the gender specific tolerance threshold for that individual's school $s(i)$. Letting R_i indicate referral and $\tau_{\theta,s(i)}$ represent the school-gender specific tolerance threshold, we have the following selection equation for $\theta \in \{m, f\}$ and standard normal error η_i:[13]

$$R_i|x_i, \theta_i = \mathbb{1}(x_i > \tau_{\theta,s(i)} + \eta_i) \tag{2}$$

4.1 Estimation Procedure

In the data, we observe a students gender $\theta \in \{m, f\}$, their school $s(i) \in 1, ...S$, and whether or not they were referred for assessment $R_i \in \{0, 1\}$. We only observe disruption, x_i, and standardized IQ score, z_i, for the set of students that are referred. Note that we also observe whether or not an individual is identified as gifted. Because this occurs only when an individual is referred, and we know the IQ score if referred, then giftedness identification given referred is known with certainty and determined by $z_i \geq 2$. Therefore, the probability we are interested in is the probability of referral given a child is gifted as this will be the same as the probability of giftedness identification given a child is gifted.

For estimation purposes, we need only focus on referral vs non referral in the likelihood function. For a student i at school $s(i)$ and gender θ_i we have the following two scenarios in the data:

- Case 1: Child is referred and both x_i and z_i are observed by the econometrician $\rightarrow (R_i = 1, x_i = x, z_i = z)$

Let $P_1(x, z, \alpha_\theta, \tau_{\theta,s})$ be defined as the likelihood of observing $(R_i = 1, x_i = x, z_i = z)$ for a child at school s and of gender θ, conditional on parameters α, τ. This can be written as a chain of conditional probabilities with a closed-form solution of observed data and model parameters:

$$\begin{aligned}
P_1(x, z, \alpha_\theta, \tau_{\theta,s}) &= Pr(R_i = 1, x_i = x, z_i = z) \\
&= Pr(R_i = 1 \mid x_i = x, z_i = z) \times Pr(x_i = x \mid z_i = z) \times Pr(z_i = z) \\
&= Pr(x_i > \tau_{\theta,s} + \eta_i) \times Pr(\alpha_{0,\theta} + \alpha_{1,\theta} * z + \alpha_{2,\theta} * z^2 + \varepsilon_i = x) \times Pr(z_i = z) \\
&= \Phi(x - \tau_{\theta,s}) \times \phi(x - \alpha_{0,\theta} + \alpha_{1,\theta} * z + \alpha_{2,\theta} * z^2) \times \phi(z)
\end{aligned}$$

[13] This η_i differs from the one in equation 1. This error allows for deviation in the student's referral from that of other students in the same school. An example could be ethnicity, family emergency, etc.

- Case 2: Child is not referred so neither x_i nor z_i is observed by the econometrician $\rightarrow (R_i = 0)$

Recall that for the set of children with $R_i = 0$, we do not observe x_i nor z_i. Thus, this likelihood is only a function of model parameters. Let $P_0(\alpha_\theta, \tau_{\theta,s})$ be defined as the likelihood of observing $(R_i = 0)$ for a child at school s and of gender θ, conditional on parameters α, τ. This can also be re-written as a conditional probability, though determining its contribution to the likelihood depends on integration over the known distribution of IQ, which we estimate via simulations:

$$\begin{aligned} P_0(\alpha_\theta, \tau_{\theta,s}) &= Pr(R_i = 0) \\ &= \int Pr(R_i = 0 \mid z_i = z_k) \times \phi(z_k) dz_k \\ &= \int Pr(x_i < \tau_{\theta,s} + \eta_i \mid z_k) \times \phi(z_k) dz_k \\ &= \int \Phi((2)^{-1/2}(\tau_{\theta,s} - \alpha_{0,\theta} - \alpha_{1,\theta} * z_k - \alpha_{2,\theta} * z_k^2)) \times \phi(z_k) dz_k \end{aligned}$$

Putting the two cases together, the likelihood of observing (R_i, x_i, z_i) for a given child at school s and gender θ is: $L(R_i, x_i, z_i \mid s, \theta, \alpha, \tau) = 1(R_i = 1) log P_1(x, z, \alpha_\theta, \tau_{\theta,s}) + 1(R_i = 0) log P_0(\alpha_\theta, \tau_{\theta,s})$. Finally, we aggregate over children in each school and across all schools in the sample and find the set of model parameters that maximize the log-likelihood below, with $N^0_{\theta,s}$ is the number of $\theta \in \{male, female\}$ children in school, s, that were not referred to psychological services for assessment:

$$\begin{aligned} \max_{\hat{\alpha},\hat{\tau}} \sum_{s=1}^{S} \sum_{i \in s} \Big[& \mathbb{1}(R_i = 1) \left[\mathbb{1}(\theta_i = m) log P_1(x, z, \alpha_m, \tau_{m,s}) + \mathbb{1}(\theta_i = f) log P_1(x, z, \alpha_f, \tau_{f,s}) \right] \\ & + N^0_{m,s} log P_0(\alpha_m, \tau_{m,s}) + N^0_{f,s} log P_0(\alpha_f, \tau_{f,s}) \Big] \end{aligned}$$

5 Results

We use the data described in section 2 to estimate the underlying parameters of the model.

We estimate the following parameters via maximum likelihood estimation: $\alpha_{0,\theta}, \alpha_{1,\theta}, \alpha_{2,\theta}$, and $\tau_{\theta,s}$ for $\theta_i \in \{\text{male, female}\}$ and school $s = 1, ..., S$.

Table 2 presents the maximum likelihood estimates for the parameters that define the rela-

Table 2: IQ-Disruption Model Estimates

	(Male)	(Female)
α_0	-0.463	-0.619
	(0.027)	(0.031)
α_1	-0.092	-0.200
	(0.023)	(0.025)
α_2	0.021	-0.005
	(0.014)	(0.018)

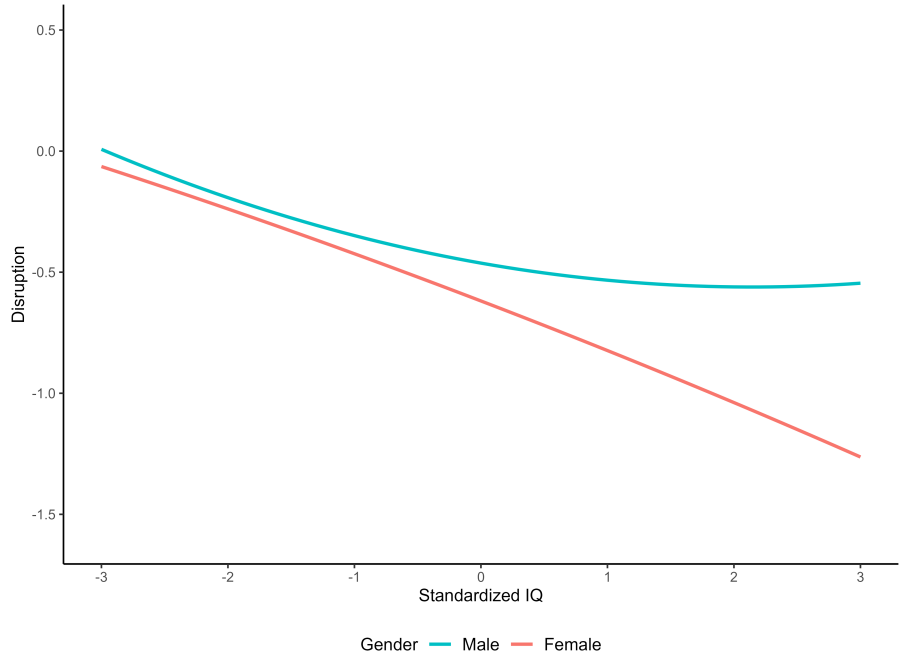
Note: This table presents the model parameter estimates for $\hat{\alpha}_{0,\theta}, \hat{\alpha}_{1,\theta}$ and $\hat{\alpha}_{2,\theta}$ for θ = male and θ= female. Standard errors based on numerical approximation of hessian in parenthesis.

tionship between disruption and IQ scores ($\alpha_{0,\theta}$, $\alpha_{1,\theta}$ and $\alpha_{2,\theta}$) and Figure 4 presents a visual of this estimated relationship. Compared to the OLS fit in Figure 3, the average level of disruption (y-axis) is lower. This is because the OLS fit is conditional on being referred to the school psychological service while the maximum-likelihood estimated relationship accounts for this selection. In other words, Figure 4 presents the IQ-disruption for the full sample of children and Figure 3 includes only those with high enough disruption to be referred.

Adjusting for selection into referral, we find that disruption levels are higher for male students than for female students at all points in the IQ distribution. We also find that disruption initially decreases with IQ for both males and females; however, while the relationship is U-shaped for male students ($\alpha_{2,\text{Male}}$ is statistically significantly different than 0), it is not for female students. This supports the claim that the giftedness gender gap can be explained by higher level of externalized disruptive behaviors at high IQ among males but not females.

One additional driver for referral to the school psychological service is the school-and gender-specific tolerance threshold, $\tau_{\theta,s}$. In Figure 5, we show the relationship between male and female tolerance threshold, by school. For all but 5 schools in our sample, we see that the male tolerance threshold is lower than the female tolerance threshold. This provides empirical evidence that teachers are on average more likely to refer male students than female students to the psychological service for assessment, even when male students and female students display the same behaviors. Put another way, the fact that tolerance thresholds are higher for female students implies that teachers or schools will allow female students to be "more disruptive" before they decide to refer them. For example, it might be that male students only have to act

Figure 4: IQ-Disruption Relationship (Maximum Likelihood Estimates)



Note: This figure plots the gender specific IQ-disruption relationship using $\alpha_{0,\theta}, \alpha_{1,\theta}$ and $\alpha_{2,\theta}$ estimates obtained from our maximum likelihood procedure. The estimated parameters and standard errors are displayed in Table 2.

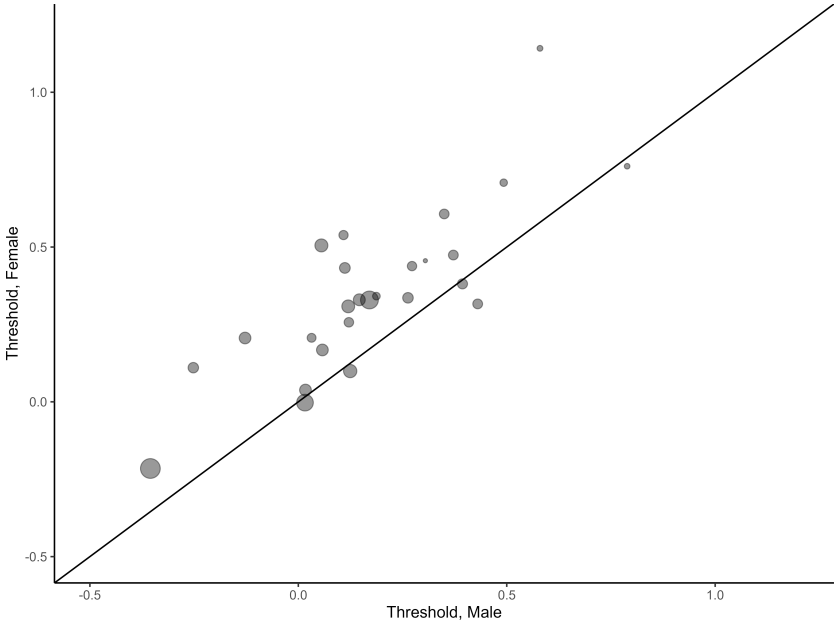
out once whereas female students might need to act out more frequently until the teacher or the school will refer them. This, in addition to the lack of U-shape in the female disruption-IQ relationship, explains why gifted female students are more likely to be under-identified when compared to their male counterparts.

5.1 Simulated Decomposition Exercises

Next, we study the relative importance of the two mechanisms in explaining the underidentification of gifted female students. We do so by examining how the probability of referral at different levels of IQ changes when male model parameters are used for female students. Recall that giftedness identification occurs when a child is referred *and* is gifted. Because all gifted students who are referred are subsequently tested and identified as gifted deterministicly, we are interested in the probability of referral at high values of IQ as this will be equal to the probability of giftedness identification at high values of IQ.

Solid lines in Figure 6 show the probability of referral (y-axis) conditional on standardized IQ score (x-axis) for male and female students based on their own model estimates, assuming they attend the median referring school. These probabilities are depicted as $P_1(\alpha_m, \tau_m)$ and

Figure 5: Male/Female Tolerance Thresholds, by School

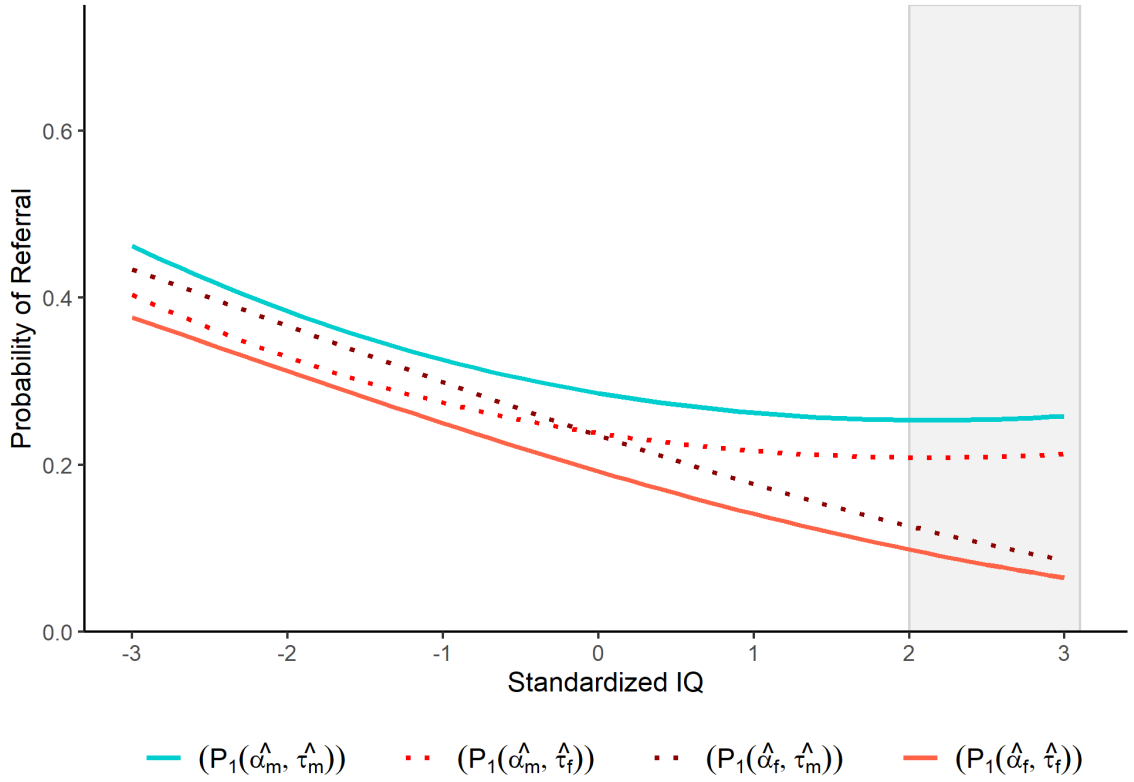


Note: This figure plots the male tolerance threshold estimate and female tolerance threshold estimate, where each dot represents a school in the sample. The size of each dot is proportional to the within-school average estimate standard error, with larger dots corresponding to higher estimate precision. 45-degree line included for comparison.

$P_1(\alpha_f, \tau_f)$. For all levels of IQ, we see that males are more likely to be referred to the school psychological service than females. At $z = 2$, corresponding to the giftedness threshold of $IQ = 130$, boys have a 25.3% probability of being referred, while the probability for girls is 9.9%. Therefore, gifted boys are at least 2.5 times more likely to be identified than gifted girls.

Dotted lines in Figure 6 show the counterfactual probabilities. We see that if teachers/schools used the same tolerance thresholds for females as they do with male students ($P_1(\alpha_f, \tau_m)$), the probability of referral would increase slightly and only approach the male rate at lower measures of IQ. However, if females kept their estimated tolerance threshold but had the same IQ-disruption relationship as males ($P_1(\alpha_m, \tau_f)$), the referral rate gap would close significantly. Specifically, for females with $IQ = 130$, the referral rate would rise to 20.8% and thus the referral rate gap would be reduced by 71%. In conclusion, this figure shows that for gifted students, the gender gap in referral (and thus giftedness identification) is mostly driven by differences in the underlying IQ-disruption relationship ($\alpha's$) rather than different tolerance thresholds (τ).

Figure 6: Counterfactual Exercise: Probability of Referral



Note: This figure plots the counterfactual exercise where $P_1(\alpha_m, \tau_m)$ and $P_1(\alpha_f, \tau_f)$ show the probability of referral conditional on standardized IQ score for male and female based on their respective model estimates, assuming they attend the school with a referral rate at the median of all estimated referral rates. $P_1(\alpha_f, \tau_m)$ shows the probability of referral for females, had they the same tolerance threshold as males. $P_1(\alpha_m, \tau_f)$ shows the probability of referral for females, had they the same IQ-disruption relationship as males.

6 Conclusions

We study the mechanisms that explain the underidentification of intellectual giftedness in female students. Our results suggest that heuristics used by teachers to refer students for psychometric assessments are gender-biased: gifted male students are more likely to be referred to the school psychological service, since they are more likely than female students to show behaviors that will make them referred, and subsequently assessed and identified as gifted. Differential referral thresholds for female students and male students can also be observed, but they are an order of magnitude less important in explaining the gender gap.

These results hold important policy conclusions. First, two-stage identification systems to identify intellectually gifted children are biased against female students. This has been suggested in the literature in the context of gender and ethnicity (Card and Giuliano, 2016a,b; Peters and Matthews, 2016). A possible solution could be to stop relying on two-stage referral

procedures for giftedness identification. A solution in this direction, albeit expensive, would be to use universal screening. Universal screening would not only be advantageous for female students, but could also reduce disadvantages for migrant and low-SES children. This has been suggested in different settings, e.g. in screening for school-readiness (Balestra, Eugster, and Liebert, 2020). Second, principals should increase the awareness of teachers to gifted students without socio-emotional problems.

The number of never-assessed gifted female students might have further implications in light of the literature on gender bias in education and in gender gaps in educational outcomes.

References

Balestra, Simone, Beatrix Eugster, and Helge Liebert. 2020. "Summer-born struggle: The effect of school starting age on health, education, and work." *Health Economics* 29:591–607.

———. 2022. "Peers with special needs: effects and policies." *Review of Economics and Statistics* 104 (3):1–17.

Balestra, Simone, Aurélien Sallin, and Stefan C Wolter. 2023. "High-Ability Influencers?: The Heterogeneous Effects of Gifted Classmates." *Journal of Human Resources* 58 (2):633–665.

Bertrand, Marianne. 2011. "New Perspectives on Gender." Elsevier, 1543–1590.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households." *The Quarterly Journal of Economics* 130 (2):571–614.

Bertrand, Marianne and Jessica Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5 (1):32–64.

Bianco, Margarita, Bryn Harris, Dorothy Garrison-Wade, and Nancy Leech. 2011. "Gifted Girls: Gender Bias in Gifted Referrals." *Roeper Review* 33 (3):170–181.

Blau, Francine D. and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55 (3):789–865.

Booij, Adam, Ferry Haan, and Erik Plug. 2016. "Enriching Students Pays Off: Evidence from an Individualized Gifted and Talented Program in Secondary Education." .

Brown, Thomas, Philipp Reichel, and Donald Quinlan. 2009. "Executive function impairments in high IQ adults with ADHD." *Journal of Attention Disorders* 13 (2):161–167.

Bui, Sa A, Steven G Craig, and Scott A Imberman. 2014. "Is gifted education a bright idea? Assessing the impact of gifted and talented programs on students." *American Economic Journal: Economic Policy* 6 (3):30–62.

Caldwell, Sydnee and Oren Danieli. 2024. "Outside Options in the Labor Market." *The Review of Economic Studies* 91 (6):3286–3315.

Calvo, Paula, Ilse Lindenlaub, and Ana Reynoso. 2024. "Marriage Market and Labor Market Sorting." *The Review of Economic Studies* .

Card, David, Eric Chyn, and Laura Giuliano. 2024. "Can Gifted Education Help Higher-Ability Boys from Disadvantaged Backgrounds?" Tech. rep., National Bureau of Economic Research.

Card, David and Laura Giuliano. 2016a. "Can tracking raise the test scores of high-ability minority students?" *American Economic Review* 106 (10):2783–2816.

———. 2016b. "Universal screening increases the representation of low-income and minority students in gifted education." *Proceedings of the National Academy of Sciences* 113 (48):13678–13683.

Carlana, Michela. 2019. "Implicit Stereotypes: Evidence from Teachers' Gender Bias." *The Quarterly Journal of Economics* 134 (3):1163–1224.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *The Quarterly Journal of Economics* 125 (3):1101–1144.

Cohen, Krystal. 2022. "Young, gifted, and black: Inequitable outcomes of gifted and talented programs." *Journal of Public & International Affairs* 8.

Cortés, Patricia and Jessica Pan. 2023. "Children and the Remaining Gender Gaps in the Labor Market." *Journal of Economic Literature* 61 (4):1359–1409.

Elder, Todd E. 2010. "The importance of relative standards in ADHD diagnoses: evidence based on exact birth dates." *Journal of health economics* 29 (5):641–656.

Elder, Todd E, David N Figlio, Scott A Imberman, and Claudia L Persico. 2021. "School segregation and racial gaps in special education identification." *Journal of Labor Economics* 39 (S1):S151–S197.

Eugster, Beatrix. 2017. "Children with special needs: long-term effects of early therapy and peer composition in schools." Swiss National Science Foundation Project, Center for Disability and Integration CDI-HSG Universität St. Gallen.

Exley, Christine L. and Judd B. Kessler. 2022. "The Gender Gap in Self-Promotion." *The Quarterly Journal of Economics* 137 (3):1345–1381.

Guénolé, Fabian, Jacqueline Louis, Christian Creveuil, Jean-Marc Baleyte, Claire Montlahuc, Pierre Fourneret, and Olivier Revol. 2013. "Behavioral profiles of clinically referred children with intellectual giftedness." *Biomed Res Int.* .

Hafenstein, Norma L, Vicki Boley, and Joi Lin. 2022. "State Policy and Funding in Gifted Education." *Gifted Child Today* 45 (4):226–234.

Harris, Angel H, Darryl V Hill, and Matthew A Lenard. 2022. "Promises, Pitfalls, and Tradeoffs in Identifying Gifted Learners: Evidence from a Curricular Experiment." .

Lavy, Victor and Yoav Goldstein. 2022. "Gifted Children Programs' Short and Long-Term Impact: Higher Education, Earnings, and the Knowledge-Economy." .

Layton, Timothy J, Michael L Barnett, Tanner R Hicks, and Anupam B Jena. 2018. "Attention deficit–hyperactivity disorder and month of school enrollment." *New England Journal of Medicine* 379 (22):2122–2130.

Lazear, Edward P. 2001. "Educational production." *The Quarterly Journal of Economics* 116 (3):777–803.

Marquardt, Kelli. 2023. "Mis (sed) diagnosis: Physician decision-making and ADHD." .

McBee, Matthew T, Scott J Peters, and Erin M Miller. 2016. "The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis." *Gifted Child Quarterly* 60 (4):258–278.

McClain, Mary-Catherine and Steven Pfeiffer. 2012. "Identification of gifted students in the United States today: A look at state definitions, policies, and practices." *Journal of Applied School Psychology* 28 (1):59–88.

Nyborg, Helmuth. 2015. "Sex differences across different racial ability levels: Theories of origin and societal consequences." *Intelligence* 52:44–62.

Persson, Petra, Xinyao Qiu, and Maya Rossin-Slater. 2021. "Family spillover effects of marginal diagnoses: The case of ADHD." .

Peters, Scott and Michael Stuart Matthews. 2016. "Gifted education research from the economists' perspective: What have we learned?" *Journal of Advanced Academics* (2):150–161.

Petersen, Jennifer. 2013. "Gender differences in identification of gifted youth and in gifted program participation: A meta-analysis." *Contemporary Educational Psychology* (4):342–348.

Rinn, A. N., R. U. Mun, and J. Hodges. 2022. "2020-2021 State of the States in Gifted Education." Tech. rep., National Association for Gifted Children and the Council of State Directors of Programs for the Gifted.

Rutigliano, Alexandre and Nikita Quarshie. 2021a. "Policy approaches and initiatives for the inclusion of gifted students in OECD countries." (262). URL `https://www.oecd-ilibrary.org/content/paper/c3f9ed87-en`.

———. 2021b. "Policy approaches and initiatives for the inclusion of gifted students in OECD countries." .

Schwandt, Hannes and Amelie Wuppermann. 2016. "The youngest get the pill: ADHD misdiagnosis in Germany, its regional correlates and international comparison." *Labour Economics* 43:72–86.

Tasca, Ilaria, Michele Guidi, Patrizia Turriziani, Giovanni Mento, and Vincenza Tarantino. 2022. "Behavioral and Socio-Emotional Disorders in Intellectual Giftedness: A Systematic Review." *Child Psychiatry Hum Dev.* .

VanTassel-Baska, Joyce. 2018. "American policy in gifted education." *Gifted Child Today* 41 (2):98–103.

Wasserman, Melanie. 2023. "Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents." *The Review of Economic Studies* 90 (3):1535–1568.

Wiswall, Matthew and Basit Zafar. 2018. "Preference for the Workplace, Investment in Human Capital, and Gender." *The Quarterly Journal of Economics* 133 (1):457–507.

A Estimation procedure

For each child i, we observe their gender θ_i, which school they attend ($s \in 1, ..., S$), and an indicator for referral $R_i \in 0, 1$. If the child is referred, we also observe standardized IQ score z_i and the level of disruptiveness x_i. Further, while gifted identification is deterministic in the model, we do confirm in the data that only referred children with IQ above XX (i.e. those with $R_i == 1$ and $z_i > 2$) are labeled as "gifted".

For estimation purposes, we need consider only the following two possible outcomes conditional on school s_i and gender θ_i:

- Case 1: Child is referred $\rightarrow (R_i = 1, x_i = x, z_i = z)$

Let $P_1(x, z, \alpha_\theta, \tau_{\theta,s})$ be defined as the likelihood of observing $(R_i = 1, x_i = x, z_i = z)$ for a child at school s and of gender θ, conditional on parameters α, τ. This can be written as a chain of conditional probabilities with an analytical formula:

$$\begin{aligned} P_1(x, z, \alpha_\theta, \tau_{\theta,s}) &= Pr(R_i = 1, x_i = x, z_i = z | s, \theta) \\ &= Pr(R_i = 1 | x_i = x, z_i = z) \times Pr(x_i = x | z_i = z) \times Pr(z_i = z) \\ &= Pr(x_i > \tau_{\theta,s} + \eta_i) \times Pr(\alpha_{0,\theta} + \alpha_{1,\theta} * z + \alpha_{2,\theta} * z^2 + \varepsilon_i = x) \times Pr(z_i = z) \\ &= \Phi(x - \tau_{\theta,s}) \times \phi(x - \alpha_{0,\theta} + \alpha_{1,\theta} * z + \alpha_{2,\theta} * z^2) \times \phi(z) \end{aligned}$$

- Case 2: Child is not referred $\rightarrow (R_i = 0)$

Recall that for the set of children with $R_i = 0$, we do not observe x_i nor z_i. Thus, this likelihood is only a function of model parameters. Let $P_0(\alpha_\theta, \tau_{\theta,s})$ be defined as the likelihood of observing $(R_i = 0)$ for a child at school s and of gender θ, conditional on parameters α, τ. This can also be re-written as a conditional probability, though determining its contribution to the likelihood depends on integration over the known disruption of IQ, which we estimate via simulations:

$$\begin{aligned} P_0(\alpha_\theta, \tau_{\theta,s}) &= Pr(R_i = 0 | s, \theta) \\ &= \int Pr(R_i = 0 | z_i = z_k) \times \phi(z_k) dz_k \\ &= \int Pr(x_i | z_k < \tau_{\theta,s} + \eta_i) \times \phi(z_k) dz_k \\ &= \int \Phi((2)^{-1/2}(\tau_{\theta,s} - \alpha_{0,\theta} - \alpha_{1,\theta} * z_k - \alpha_{2,\theta} * z_k^2)) \times \phi(z_k) dz_k \end{aligned}$$

Putting the two cases together, the likelihood of observing (R_i, x_i, z_i) for a given child at school s and gender θ is: $L(R_i, x_i, z_i|s, \theta, \alpha, \tau) = 1(R_i = 1)logP_1(x, z, \alpha_\theta, \tau_{\theta,s}) + 1(R_i = 0)logP_0(\alpha_\theta, \tau_{\theta,s})$. Finally, we aggregate over children in each school and across all schools in the sample and find the set of model parameters that maximize the log-likelihood below, with $N^0_{\theta,s}$ is the number of $\theta \in \{male, female\}$ children in school, s, that were not referred to psychological services for assessment:

$$\max_{\hat{\alpha},\hat{\tau}} \sum_{s=1}^{S} \sum_{i \in s} \Big[\mathbb{1}(R_i = 1) \left[\mathbb{1}(\theta_i = m) logP_1(x, z, \alpha_m, \tau_{m,s}) + \mathbb{1}(\theta_i = f) logP_1(x, z, \alpha_f, \tau_{f,s}) \right] + N^0_{m,s} logP_0(\alpha_m, \tau_{m,s}) + N^0_{f,s} logP_0(\alpha_f, \tau_{f,s}) \Big]$$

B Alternative IQ distributions

The main assumption underlying the model is that males and females have the same underlying IQ distribution. Another hypothesis is that there exists a gender difference in IQ, with males having a higher mean and higher variance than females (Nyborg, 2015). As a robustness check, we compare our main estimates to one that assumes gender-specific IQ distributions, based on values obtained in Nyborg (2015). Specifically, we take the gender-specific IQ mean and variance from Table 1a (for males) and 1b (for females). That is, $IQ|_{\theta_i=male} \sim N(103.19, 15.57^2)$ and $IQ|_{\theta_i=female} \sim N(102.18, 13.55^2)$.

Estimation follows from above, with differences being the gender-specific standardization of observed IQ for referred students. While results are qualitatively similar to our preferred specification, there are some differences. For example, with gender-specific IQ, we see that females have a more pronounced quadratic relationship between IQ and disruption in the tails.

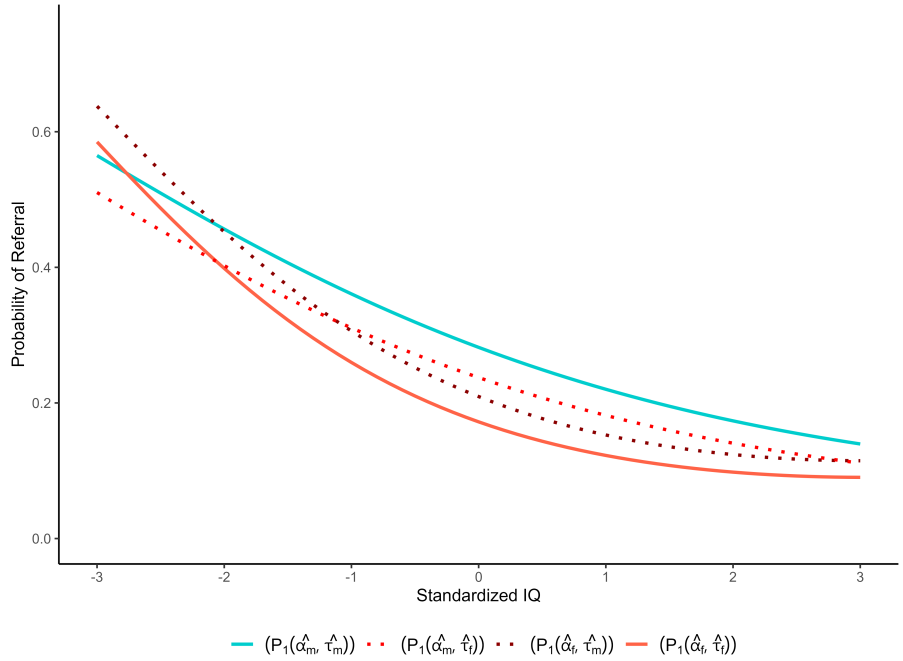
However, both models have similar interpretations for children at the margin of giftedness identification. In both scenarios, male students are more likely to be referred than female students, and this difference is mainly driven by differences in the IQ-disruption relationship rather than differences in referral tolerance thresholds. We also note that the model assuming equal IQ distributions fits the data best in terms of higher maximum likelihood values.

Table 3: Alternative IQ Distributions: IQ-Disruption Model Estimates

	(Male)	(Female)
α_0	-0.496	-0.727
	(0.028)	(0.031)
α_1	-0.208	-0.259
	(0.023)	(0.024)
α_2	0.013	0.043
	(0.015)	(0.015)

Note: This table presents the model parameter estimates for $\hat{\alpha}_{0,\theta}, \hat{\alpha}_{1,\theta}$ and $\hat{\alpha}_{2,\theta}$ for θ = male and θ= female, under the assumption of sex-specific IQ distributions. Standard errors based on numerical approximation of hessian in parenthesis.

Figure 7: Alternative IQ Distributions: Counterfactual Exercise



C Additional tables and figures

Table 4: Summary statistics for reason for referral

	Average	Min.	Max.	Sd.
Students, conditional on referral ($N = 4,002$)				
Referred for social and emotional problems	0.24	0	1	0.43
Referred for learning disabilities	0.92	0	1	0.27
Referred for other reasons	0.14	0	1	0.35

Notes: This table reports summary statistics about our reasons for referral in our sample. All students born between 1997 and 2000 and referred to the School Psychological Services (SPS) in the Swiss Canton of St. Gallen are included in the sample. 104 distinct school-cohort. *Source*: Ministry of Education, Canton of St. Gallen

Figure 8: IQ distribution for both genders in restricted sample

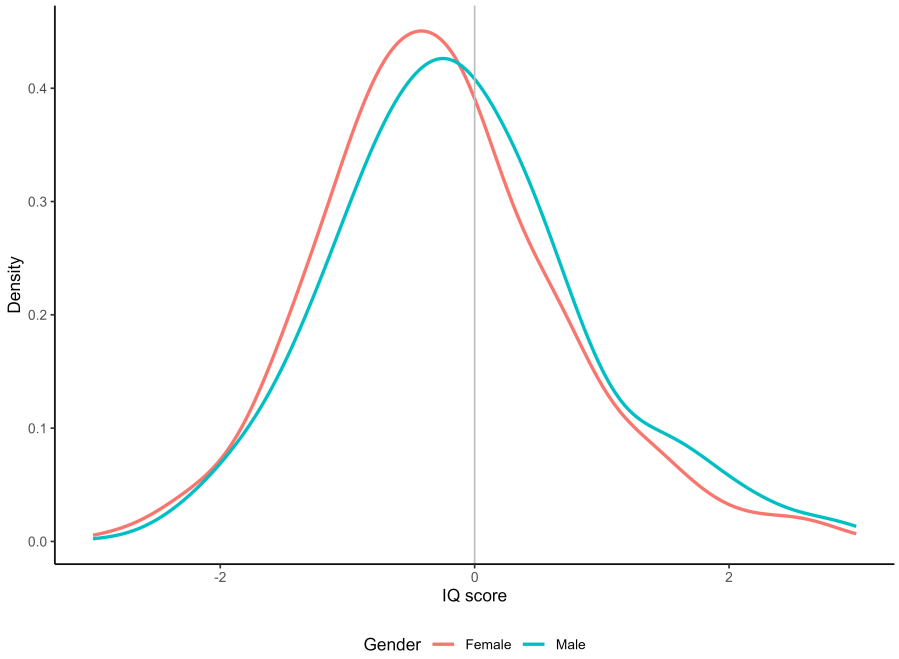


Figure 9: Distribution of our "disruption" variable for both genders in restricted sample

