

Working Papers

Missing Stars? Quantifying the Gender Gap in the Assessment of Gifted Students

Beatrix Eugster, Kelli Marquardt, Aurélien Sallin

Imprint:

CESifo Working Papers

ISSN 2364-1428 (digital)

Publisher and distributor: Munich Society for the Promotion
of Economic Research - CESifo GmbH

Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740

Email office@cesifo.de
<https://www.cesifo.org>

Editor: Clemens Fuest

An electronic version of the paper may be downloaded free of charge

- from the CESifo website: www.ifo.de/en/cesifo/publications/cesifo-working-papers
- from the SSRN website: www.ssrn.com/index.cfm/en/cesifo/
- from the RePEc website: <https://ideas.repec.org/s/ces/ceswps.html>

Missing stars?

Quantifying the gender gap in the assessment of gifted students*

Beatrix Eugster[†], Kelli Marquardt[‡], Aurélien Sallin[§]

January 2026

Abstract

Female students are less likely to be identified as intellectually gifted than male students when the identification process follows a two-stage procedure. We posit and quantify two mechanisms that explain this gender gap. First, at any level of IQ, but especially at high levels of IQ, male students are more likely nominated for assessment due to higher salience of externalizing behaviors. This difference accounts for roughly 70% of the gender gap in giftedness identification. Second, the remainder of the gap is explained by the use of lower referral thresholds for male students, all else equal. We discuss policy implications.

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

JEL Classification: I21, I24, J16

*We are grateful to Nora Bearth, Andrea Burro, Samuel Jong Kang, David Preinerstorfer, Fanny Puljic, Camille Terrier, Stefan Wolter, and participants of seminars at the University of St. Gallen and the University of Bergamo for their helpful comments. We thank Taegan Mullane for excellent research assistance. 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. This study is based on observational/administrative data, which was initially acquired for the project “Children with Special Needs” financed by the Swiss National Science Foundation (grant 176381). The analysis is therefore exempt from pre-registration.

[†]University of St.Gallen, CESifo, CEPR. Corresponding author: beatrix.eugster@unisg.ch

[‡]Federal Reserve Bank of Chicago

[§]University of St.Gallen, SWICA.

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), 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 academic achievement and graduation rates, there is a persistent gender gap in identifying gifted children, i.e., children with significantly higher intellectual ability than average. This is particularly true for disadvantaging gifted female students (Balestra, Sallin, and Wolter, 2023; Petersen, 2013). The underidentification of intellectually gifted female students may limit their access to adapted education programs and associated benefits, which could contribute to the persistence of long-term gender disparities in the economy.

The underidentification of female students (or minority students more generally) in gifted programs arises from features of how students are screened and selected into gifted education. In many settings, 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. Moreover, the 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 giftedness (e.g., classroom disruption).

In two-step identification processes, children are first referred to an assessor from parents and teachers through a nomination/referral 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 underidentifi-

cation 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 psychological service over four 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 model that allows us to decompose the two mechanisms that contribute to the gender gap in giftedness identification. 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.¹

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. In particular, we document the mechanisms that contribute to

¹The model and estimation strategies are in spirit of Marquardt (2025), which quantifies the mechanisms leading to gender differences in diagnosis rates for Attention Deficit Hyperactivity Disorder (ADHD).

²It has been argued that teacher referral is more prone to favor male students over female students in giftedness identification, while (especially at lower grade) grade point averages are more likely to flag female students as gifted (Petersen, 2013).

the disparity in early customized education programs for female gifted students. Such disparities have important long run economic implications given that most education systems within the 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

Further, this paper refines our understanding of the limitations inherent in two-stage assessment processes commonly used in education and behavioral health. Such processes have 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 students for special education services, 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). Marquardt (2025) studies the gender gap in ADHD diagnosis using a similar two-stage model as we describe in this paper. She finds that physician use of gender-specific diagnostic thresholds is the primary driver of the gender gap in the identification of ADHD, and posits that gender differences in externalizing vs internalizing behaviors may rationalize this result. Our paper contributes to this literature by illustrating how differences in classroom behaviors between males and females may 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.

³In the United States, there is no federal mandate related to gifted identification and services, which results in large variation in gifted education across states and even within districts in a single state. Despite this, Gifted and Talented (GT) education programs in the US serve more than 3 million public school students (Harris, Hill, and Lenard, 2022).

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

Although there has been significant progress in reducing gender gaps in academic achievement 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 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.

2 Background, Setting, and Data

2.1 Background: giftedness assessment around the world

Intellectually gifted children 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 15.⁶ An IQ score of 130 and above (two standard deviations above the mean) is used to diagnose individuals as gifted (McClain and Pfeiffer, 2012). This means that around 2.5% of the population is gifted as defined by IQ.⁷

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 US for example, each state has 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, 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). Gifted education also varies significantly across states, and even

⁵US 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>

⁶Note that IQ tests are normed to be gender-neutral. However, girls often perform better in subtests on verbal processing and attention control while boys outperform girls in spatial manipulation tasks (Giofrè et al., 2024).

⁷While our theoretical analysis in this paper primarily relies on the traditional IQ based definition of giftedness, we note that over recent decades many US States and European countries have broadened their definitions beyond intellectual ability alone. These expanded frameworks often incorporate additional dimensions such as academic achievement, creativity and creative thinking, artistic ability and leadership (Rinn, Mun, and Hodges, 2022). Gender differences in these characteristics and/or in the salience of such characteristics could also exacerbate disparities in giftedness identification and is an important area of future research. Note that in our empirical application, children can be identified as gifted based on the broader definition.

across school districts within a single states. There is no federal mandate related to gifted identification and services.

That said, the most common processes for identifying gifted students often rely on a two-stage identification procedure. In the first stage, parents or teachers refer the child to an assessor (nomination or referral stage); in the second stage, a psychologist assesses the child’s giftedness (for a detailed discussion of the U.S. process, see McBee, Peters, and Miller, 2016). In the US, referral mechanisms differ between states: in 2011, 16 (17) states mandated the use of intelligence (achievement) tests, 13 states required nominations (teacher and/or parent referrals), and 9 states required a teacher-completed behavior rating scale (McClain and Pfeiffer, 2012). In Europe, procedures also vary but often rely on nominations (Rutigliano and Quarshie, 2021b).

Because referral is the first step in the identification process, disparities in referral rates directly affect the likelihood of being identified as gifted. This referral-based process has been shown to miss many qualified students, particularly from minority and economically disadvantaged groups (Card and Giuliano, 2016b; McBee, Peters, and Miller, 2016). Inequities in gifted assessment and access have been a source of controversy, with concerns that they widen achievement gaps, especially along racial lines (Cohen, 2022).

Gender disparities in the referral rate for giftedness assessment lead to more male students being assessed for giftedness. Male students are referred for psychological assessment more often, in part because they tend to display more externalizing behaviors, whereas female students are more likely to exhibit internalizing behaviors (Bartels et al., 2018; Crijnen, Achenbach, and Verhulst, 1997; Vergunst et al., 2023; Verhulst et al., 2003). In fact, classrooms with a higher proportion of female peers have been found to experience less disruption and violence (Lavy and Schlosser, 2011). In the classroom, such disruptive behaviors draw teachers’ attention and time, increasing the chances of referral (Lazear, 2001).⁸ In our data, we show that male children have higher measures of disruption on average, with the gender gap largest at the upper end of the IQ distribution. This empirical finding is consistent with the hypothesis that giftedness is more often correlated with disruptive behaviors (Brown, Reichel, and Quinlan, 2009; Guénolé et al., 2013; Tasca et al., 2022).⁹

⁸Disruption does not have to be negative. For example, students asking very complicated questions can be as disruptive as students misbehaving.

⁹The gender disparity in referral mirrors patterns documented in ADHD assessment (Marquardt, 2025). Girls and women with ADHD more often present with inattention and internalizing problems, while boys and men show greater hyperactive-impulsive symptoms and externalizing problems (Hinshaw et al., 2022; Mowlem et al.,

2.2 Setting: giftedness assessment in St. Gallen, Switzerland

In the Swiss setting, students with special educational needs, including gifted students, are identified and assessed during a two-stage identification processes. First, students are nominated to the School Psychological Service (SPS), a centralized “gatekeeper” that subsequently identifies diagnoses, assesses special educational needs, provides psychological counseling, and refers students for further treatment. The SPS in the Canton of St. Gallen is centralized at the regional level and all regional offices serve several 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.

For most students (about nine out of ten), referral is done by the teacher, but some referrals are also submitted by parents or the child’s doctor. Referral arises for a variety of reasons, including behavioral and socio-emotional issues, learning disabilities, family problems, and more. Of the 30 percent of students referred to the SPS, around three-fourths are referred for learning disabilities. Another fourth is referred for socio-emotional reasons, potentially combined with learning impairments (Balestra, Eugster, and Liebert, 2022). Referrals for other reasons are rare.

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. In case of a persistent problematic school or learning situation, a student can be referred to the SPS repeatedly to reassess potential measures and therapies.

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.

2019). Clinicians may overlook ADHD in females due to less overt symptoms and compensatory strategies, and girls with ADHD may behave more adaptively in school settings than boys (Derks, Hudziak, and Boomsma, 2007). These differences reduce the likelihood of referral for girls, even among those with comparable levels of impairment.

2.3 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 SPS of the Canton of St. Gallen. For further background information on the data and the setting, see Balestra, Eugster, and Liebert (2020, 2022).¹⁰ 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 at referral, school of referral, and number of contacts with the SPS. In what follows, we will use the number of contacts with the SPS as a proxy for the student’s level of classroom “disruptiveness”. On average, referred students have 7.16 contacts with the SPS (7.7 on average for male students, 6.5 for female students). Although this is arguably a crude measure, it is validated by discussions with the president of the SPS, who confirmed that more disruptive students are more likely to visit the SPS repeatedly for situation assessments and assistance.¹¹

Most importantly, we observe a measure of IQ for most of the students referred to the SPS.¹² We classify a child as gifted if the IQ score is equal to or above 130, or if the psycholo-

¹⁰The 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”).

¹¹We do not find evidence that, conditional on the same number of SPS contacts, the intensity and/or inferred severity per visit differs between male and female students. In Appendix B, we show that this measure is homogeneous across gender.

¹²Intelligence tests conducted in St. Gallen are Snijders–Oomen nonverbal intelligence tests (SON), Kaufman Assessment Battery for Children I and II (K-ABC and K-ABCII), the Wechsler Intelligence Scale for Children (WISC-Hawik), Raven’s Progressive Matrices (Raven), Kramer-Test, and Culture Fair Intelligence Test (CFT). When different tests are conducted for the same student, we use a composite score of the different results, as documented 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.

Table 1: Summary statistics

	Average	Min.	Max.	Sd.	<i>N</i>
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 ($N = 4,000$)					
Female	0.44	0	1	0.50	
Age at referral	8.68	3.94	15.45	2.09	
Number of visits (disruption)	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
Male Students, conditional on referral to the SPS ($N = 2,247$)					
Age at referral	8.68	4.04	15.45	2.12	
Number of visits (disruption)	7.67	1	85	6.82	
Tested for IQ	0.73	0	1	0.44	
IQ	98.59	42	145	12.57	
Identified as gifted	0.05	0	1	0.22	116
Female Students, conditional on referral to the SPS ($N = 1,753$)					
Age at referral	8.68	3.94	15.1	2.05	
Number of visits (disruption)	6.51	1	42	5.26	
Tested for IQ	0.71	0	1	0.45	
IQ	96.54	61	138	11.76	
Identified as gifted	0.03	0	1	0.16	48

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

gist 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.¹³ 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. In total, 164 students were identified as gifted, of which 116 were male and 48 were female. The mechanisms contributing to this gender gap are what we study in the remainder of this paper.

¹³We 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).

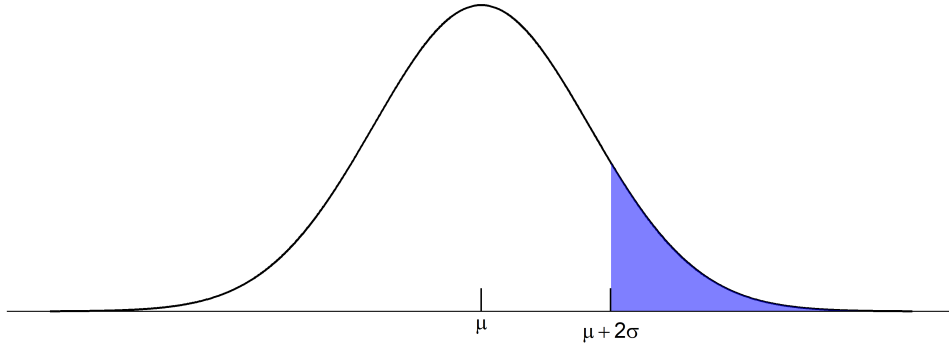
3 Theoretical Framework and Motivation

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.

3.1 Theoretical framework

To start, let v_i be an individual's IQ score that follows an arbitrary distribution with mean μ and standard deviation σ , that is, $v_i \sim F(\mu, \sigma^2)$. Let G_i be an indicator of whether the individual is considered “intellectually gifted” according to the definition of having an IQ score at least two standard deviations above the mean. Formally, $G_i = \mathbb{1}(v_i \geq \mu + 2\sigma)$. The share of the population with $G_i = 1$ corresponds to the shaded blue region in Figure 1.

Figure 1: IQ distribution and giftedness definition

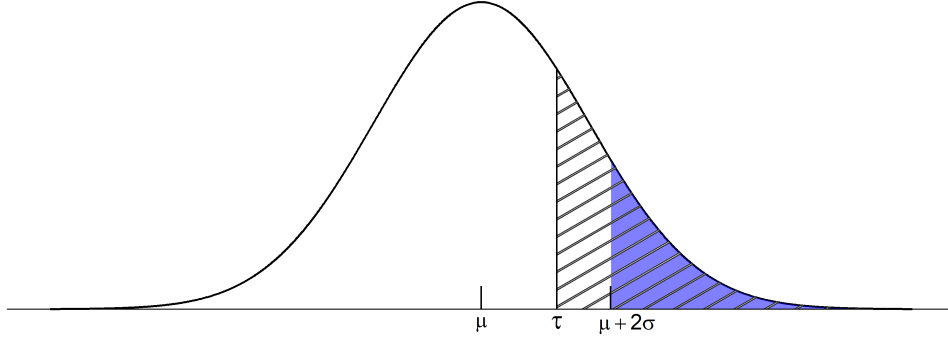


Note: IQ distribution with mean μ and standard deviation σ . The shaded blue region is the mass of gifted students.

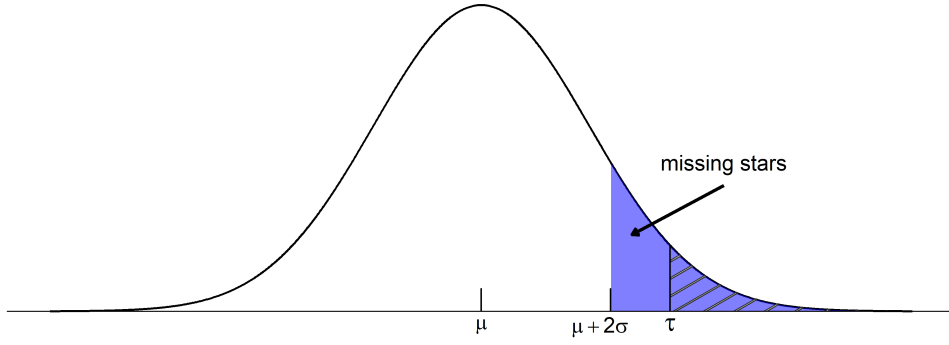
Because everyone who is referred by a teacher is also given an IQ test and an assessment of giftedness, the probability that an individual is identified as gifted is equal to the probability that they are referred and gifted. Formally, let R_i be an indicator for whether the student is referred and let 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: $Pr(IG_i = 1) = Pr(R_i = 1 \cap G_i = 1) = Pr(R_i = 1 | G_i = 1) \times Pr(G_i = 1)$.

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 $R_i = \mathbb{1}(v_i > \tau)$, and so long as $\tau < \mu + 2\sigma$, we would have $Pr(IG_i) = 1 \times Pr(G_i = 1)$ so all gifted students will be appropriately identified

Figure 2: Referral thresholds



(a) Low threshold: $\tau < \mu + 2\sigma$



(b) High threshold: $\tau > \mu + 2\sigma$

Note: IQ distribution with mean μ and standard deviation σ . The shaded blue region is the mass of gifted students. τ denotes the referral threshold.

as such (corresponding to subfigure 2a).¹⁴ 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, underidentification or “missing stars” occurs when $\tau > \mu + 2\sigma$ so $Pr(R_i = 1 | G_i = 1) < 1$ and thus $Pr(IG_i = 1) < 1 \times Pr(G_i = 1)$. The share of missing starts is denoted by the blue area without stripes in subfigure 2b.

In this simple scenario, note that the probability 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, $Pr(IG_i = 1) = fn(\tau, F(\mu, \sigma))$. 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, $F(\mu, \sigma)$.

Assumption 1: males (m) and females (f) have identical IQ distributions such that $F(\mu_m, \sigma_m^2) = F(\mu_f, \sigma_f^2)$.

¹⁴This 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.

Under assumption 1, the gender gap in gifted identification would be entirely driven by differences in τ such that $\tau_m \neq \tau_f$.¹⁵ In other words, in the situation where IQ was observed *and* used as the referral mechanism, then the 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 such behaviors, which we call x_i , are likely a function of an individual's IQ such that $x_i = g(v_i)$.

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, $Pr(IG_i = 1) = Pr(R_i = 1 | G_i = 1) \times Pr(G_i = 1)$, but now $Pr(R_i = 1 | G_i = 1) = Pr(x_i > \tilde{\tau} | v_i > \mu + 2\sigma) = Pr(g(v_i) > \tilde{\tau} | v_i > \mu + 2\sigma)$. So, in addition to referral thresholds ($\tilde{\tau}$) and IQ distribution parameters (μ and σ), giftedness identification will also depend on the relationship between the child's IQ, (i.e., v_i), and their level of disruptive behaviors (i.e., x_i), via the function $g(\cdot)$. Mathematically, $Pr(IG_i = 1) = fn(g(\cdot), \tilde{\tau}, F(\mu, \sigma))$.

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 $\tilde{\tau}_m \neq \tilde{\tau}_f$). In addition, it is also plausible that there are gender differences in the IQ-disruption relationship such that $g_m(\cdot) \neq g_f(\cdot)$. 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 cause differences in referral rates and giftedness identification above and beyond differences in teacher referral thresholds. In the next section, we present motivating facts from the data that help us link this conceptual framework to an empirical model for estimation.

¹⁵We start with Assumption 1, which is both easier to model and easier to interpret. This assumption is also plausible in light of the existing literature, as gender differences in intelligence have been debated extensively and remain contentious. First, average IQ scores for females and males have consistently been found to be very similar across large-scale studies and meta-analyses (Deary et al., 2007, 2003; Hyde et al., 2008; Hyde, 2005; Nyborg, 2015; Strand, Deary, and Smith, 2006). Second, following the "greater male variability hypothesis", studies document differences in variability, with slightly higher variance among males, but these differences are generally small and often sensitive to test design and sample composition. We account for differences in variability in one of our robustness checks. Third, performance differs by gender across specific cognitive domains (for example, verbal versus spatial abilities), yet these differences tend to offset one another in aggregate intelligence scores (Giofrè et al., 2024). Finally, most standardized intelligence tests are explicitly constructed and normed to eliminate overall mean differences between females and males (Neisser et al., 1996). Taken together, these findings support treating male and female IQ distributions as similar in expectation. Nevertheless, we explicitly address the issue of differential variance in a robustness analysis.

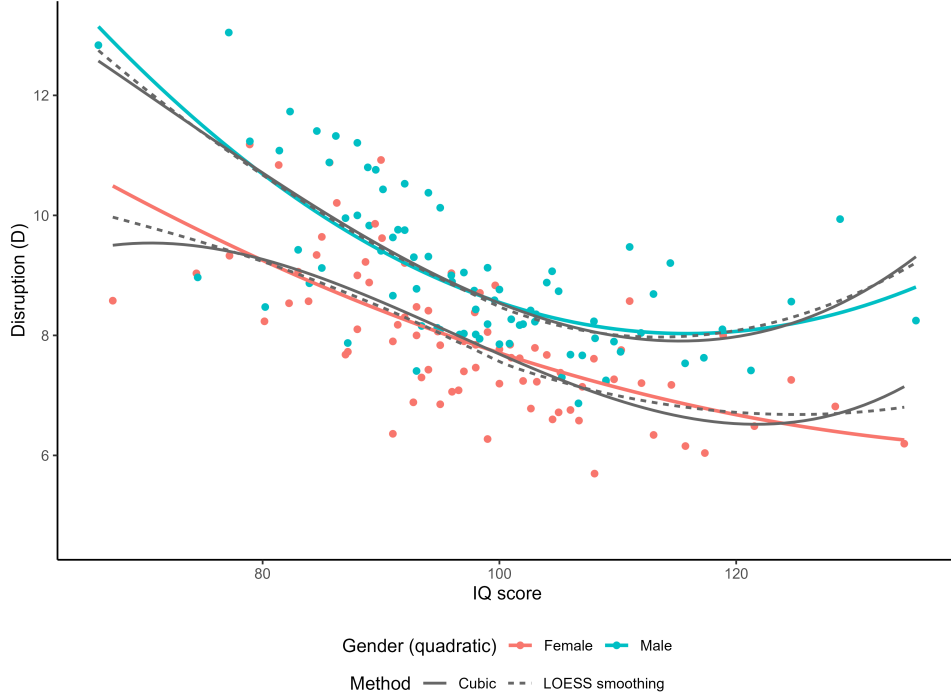
3.2 Motivating facts

We present three facts on gender differences in giftedness identification 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. Therefore, many female students 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 average referral rate for male students is 34%, while the average referral rate for female students is only 26%.

Second, we observe substantial variation in referral rates between 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). Recall that, 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 SPS. We observe that students with lower IQ scores display, on average, more severe cases and that the relationship between disruption and measured IQ is negative until a certain point. This 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.

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. 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 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.

4 Empirical Model and Estimation

We now turn to an empirical model that we can estimate, which we then use in a simulation exercise 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 3.1 in a way that is consistent with the motivating facts presented in Section 3.2. We account for both gender differences in the disruption-IQ relationship and variation in referral rates. The setup and estimation of the model closely follow that in Marquardt (2025), 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 $z_i \geq 2$ (i.e. are two standard deviations above the mean). Following assumption (1), we restrict the underlying IQ distribution to be the same for male and female students. In Section 5.2.2 and Appendix ??, we show that our main conclusions do not change when we allow for gender-specific IQ distributions.

Let x_i be our measure of classroom disruption, which we define as the number of visits to the SPS conditional on referral. For estimation purposes, we rescale the disruption measure shown in Figure 3 to be between 0 and 1.¹⁶ 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, $g(\cdot)$, as quadratic. We allow for the function parameters to differ by gender. Specifically, for $\theta \in \{m, f\}$ and standard normal error ε_i we model the relationship as in equation (1). The error ε_i captures factors that affect observed classroom behavior but are orthogonal to underlying cognitive ability and IQ (e.g., family or home stressors, temporary illness, cultural behaviors, etc.).

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

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 thresholds to differ based on the gender of the student, we also allow them to vary across schools. Note that this variation also helps us identify the α ’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.

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 .¹⁷

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

¹⁶Specifically, we rescale our measure of disruption between 0 and 1 by first referral year to account for potential changes in SPS practices.

¹⁷Note that η_i differs from ε_i in equation 1. While ε_i allows for variation in the relationship between IQ and disruption, η_i allows for variation in student’s referral propensity for a given level of disruption. Examples could be individual student-teacher relationships, behavior salience, family emergency, etc.

4.1 Estimation Procedure

In the data, we observe a student's gender $\theta_i \in \{m, f\}$, their school $s(i) \in \{1, \dots, S\}$, and whether or not they were referred for assessment $R_i \in \{0, 1\}$. For the set of students with $R_i = 1$, we also observe disruption, x_i , and standardized IQ score, z_i . We also observe whether a student is identified as gifted; however, this outcome is deterministic. Because identification occurs only for referred students and the IQ score is observed upon referral, gifted identification conditional on referral is known with certainty and is determined by the threshold $z_i \geq 2$. Therefore, the probability we are interested in modeling 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.

Thus, for estimation purposes, we need only focus on the following two cases 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. Note that the third line follows from equation (2) and (1) and the last line follows from functional form assumptions of IQ and iid error terms.

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

- Case 2: Child is not referred 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 traditional numerical approximations:

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

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: $L(R_i, x_i, z_i \mid \alpha_\theta, \tau_{\theta,s}) = \mathbb{1}(R_i = 1) \log P_1(x, z, \alpha_\theta, \tau_{\theta,s}) + \mathbb{1}(R_i = 0) \log P_0(\alpha_\theta, \tau_{\theta,s})$. 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 $N_{\theta,s}^0$ being the number of $\theta \in \{male, female\}$ children in school, s , that were not referred to psychological services for assessment:

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

5 Results

Using the data described in Section 2, we estimate the underlying parameters of the model via maximum likelihood estimation. Specifically, we obtain estimates of the following parameters: $\alpha_{0,\theta}$, $\alpha_{1,\theta}$, $\alpha_{2,\theta}$, and $\tau_{\theta,s}$ for $\theta_i \in \{male, female\}$ and school $s = 1, \dots, S$.

We start by presenting the maximum likelihood estimates for the parameters that define the relationship between disruption and IQ scores: $(\alpha_{0,\theta}, \alpha_{1,\theta}$ and $\alpha_{2,\theta})$. Table 2 presents these point estimates and standard errors. Figure 4 provides a visual of this estimated relationship. In comparing to the fit in Figure 3, note that not only are the scales different (because we standardize the y-axis value for estimation), but the fit in Figure 3 is based on those in the selected (referred) sample whereas the fit in Figure 4 is the estimated relationship for the full sample. In other words, Figure 4 shows what the IQ-disruption relationship would look like if

both measures were observed for all students and not just the referred sample of students.

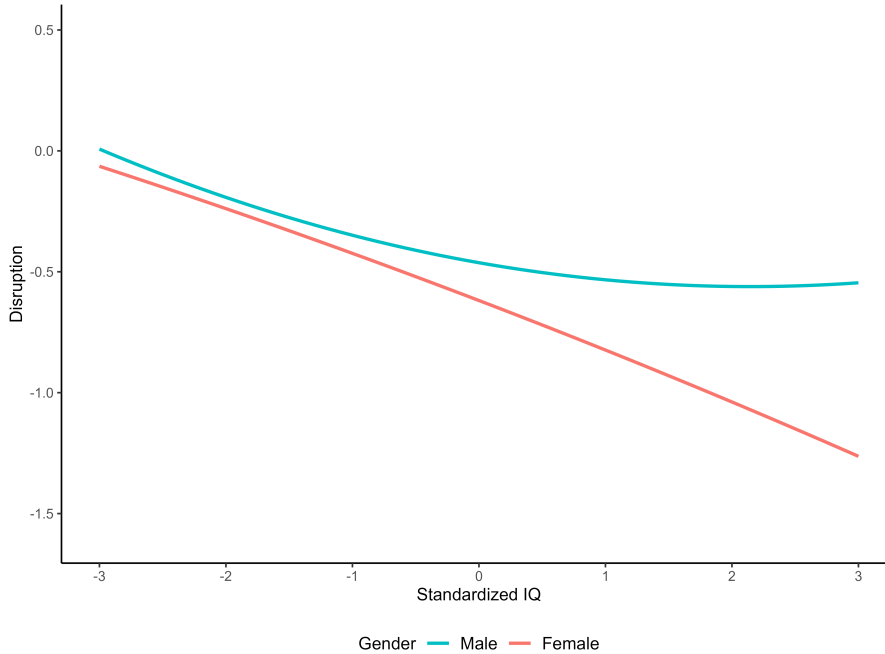
Table 2: IQ-Disruption Model Estimates

	$\theta=\text{Male}$	$\theta=\text{Female}$
$\alpha_{0,\theta}$	-0.463 (0.027)	-0.619 (0.031)
$\alpha_{1,\theta}$	-0.092 (0.023)	-0.200 (0.025)
$\alpha_{2,\theta}$	0.021 (0.014)	-0.005 (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 $\theta = \text{male}$ and $\theta = \text{female}$. Standard errors based on numerical approximation of hessian in parenthesis.

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 ($\hat{\alpha}_{2,\text{Male}}$ is statistically significantly different from 0), it is not for female students. This supports the claim that the giftedness gender gap can be explained by higher levels of externalized disruptive behaviors at high IQ among males but not females.

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

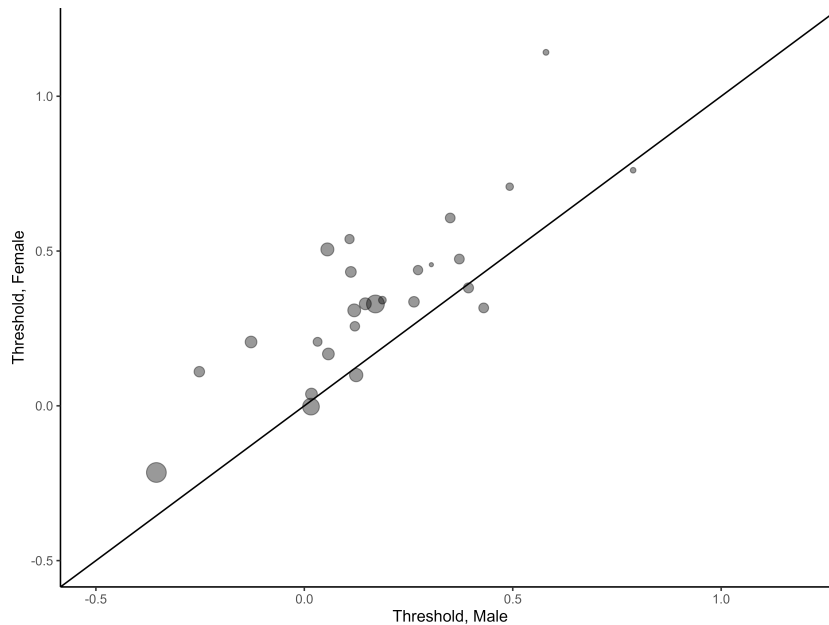


Note: This figure plots the gender specific IQ-disruption relationship using $\alpha_{0,\theta}$, $\alpha_{1,\theta}$ and $\alpha_{2,\theta}$ estimates obtained from our maximum likelihood procedure. The estimated parameters and standard errors are displayed in Table 2.

Next, we present the resulting estimates for the school-based and gender-specific tolerance

threshold, $\tau_{\theta,s}$ which is an additional mechanism leading to differences in referral and giftedness identification. 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 level of disruption. 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 out once whereas female students might need to act out more frequently until the teacher or the school will refer them. One explanation is that the externalities of the disruption may differ by gender (e.g., male classroom disruption may be more likely to have negative spillovers to peers whereas female disruption may have more internalizing consequences). 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.

Figure 5: Male/Female Tolerance Thresholds, by School



Note: This figure plots the male tolerance threshold estimate ($\hat{\tau}_{m,s}$) and female tolerance threshold estimate ($\hat{\tau}_{f,s}$), where each dot represents a school (s) in the sample. The size of each dot is proportional to the inverse within-school average estimate standard error, with larger dots corresponding to higher estimate precision. 45-degree line included for comparison.

5.1 Simulated Decomposition Exercises

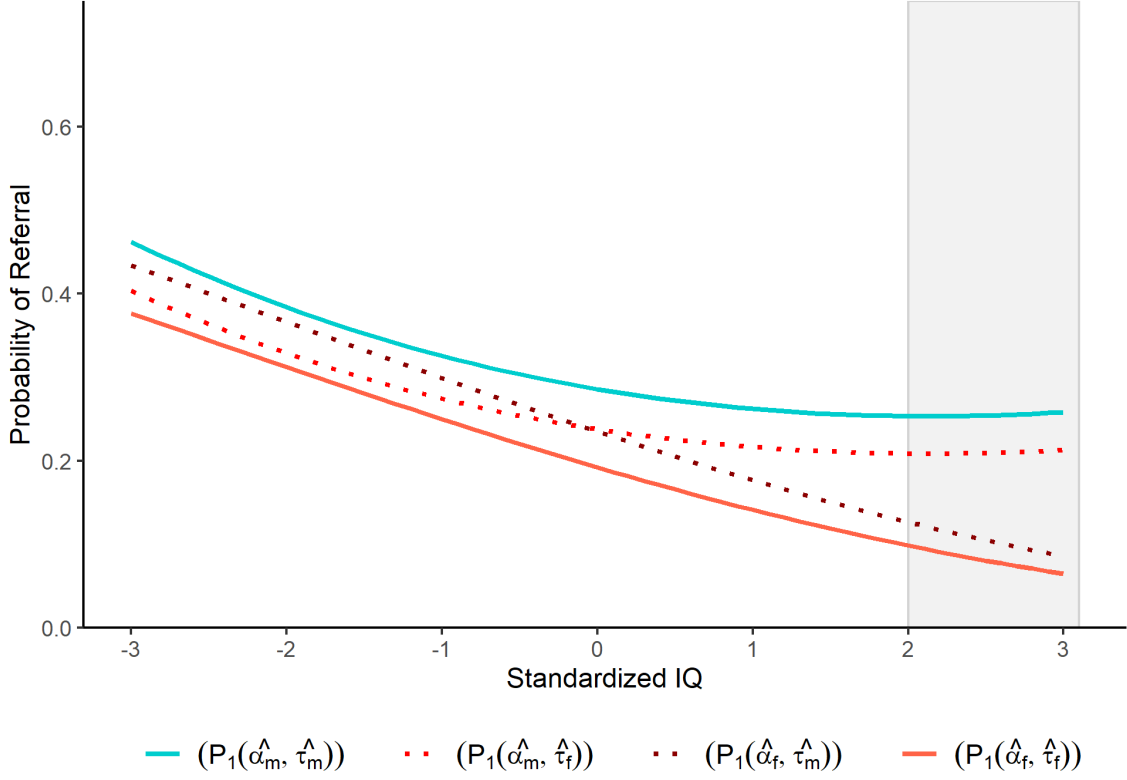
The main benefit of defining and estimating the structural model is to then use it to quantify the relative importance of the two mechanisms in explaining the underidentification of gifted female students. We do so by examining how the probability of referral at different levels of IQ restricting gender differences in model parameters. 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 deterministically, we are interested in the probability of referral at high values of IQ as this will be equal to the probability of giftedness identification.

Figure 6 provides a visualizing of our decomposition exercise. The solid lines in 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 $P_1(\hat{\alpha}_m, \hat{\tau}_m)$ and $P_1(\hat{\alpha}_f, \hat{\tau}_f)$. 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 $IQ = 130$, males have a 25.3% probability of being referred, while the probability for females is only 9.9%. Therefore, gifted males are at least 2.5 times more likely to be identified than gifted females.

Dotted lines in Figure 6 show our counterfactual probabilities. First, we ask what the referral probability would be for females if teachers/schools used the same tolerance thresholds for females as they do with male students. This is depicted by the dotted line, $(P_1(\hat{\alpha}_f, \hat{\tau}_m))$. We see that the probability of referral would increase slightly and approach the male rate at lower measures of IQ, but not at higher IQs. Thus, while this channel does contribute to the gender gap in referral, especially at low levels of IQ, it is not the driving mechanism leading to the gender gap in giftedness identification.

Next, we ask how the referral probability would change for females if they kept their estimated tolerance threshold but instead had the same IQ-disruption relationship as males. This is depicted by the dotted line, $(P_1(\hat{\alpha}_m, \hat{\tau}_f))$. Here, we see that using α_m instead of α_f closes the referral rate gap significantly at high level of IQ. 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 (α) rather than different tolerance thresholds (τ).

Figure 6: Counterfactual Exercise: Probability of Referral



Note: This figure plots the counterfactual exercise where $P_1(\hat{\alpha}_m, \hat{\tau}_m)$ and $P_1(\hat{\alpha}_f, \hat{\tau}_f)$ 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. $P_1(\hat{\alpha}_f, \hat{\tau}_m)$ shows the probability of referral for females, had they the same tolerance threshold as males. $P_1(\hat{\alpha}_m, \hat{\tau}_f)$ shows the probability of referral for females, had they the same IQ-disruption relationship as males.

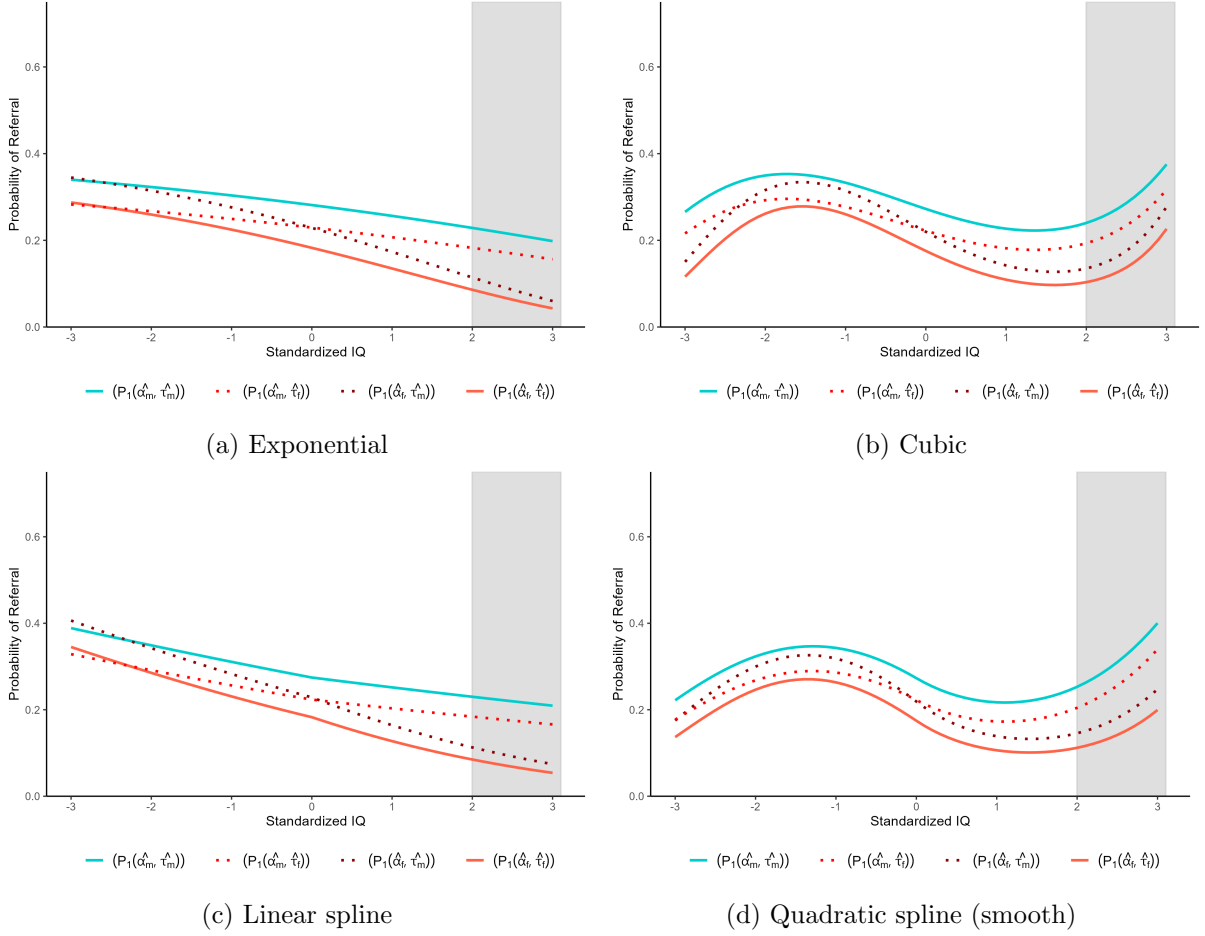
5.2 Validity

5.2.1 Functional Form

The IQ-disruption relationship as estimated above is assumed to be quadratic. Since interest is in the tail of the distribution, this could be a restrictive assumption. We therefore test the robustness of our decomposition results w.r.t. different functional form assumptions. Specifically, we re-estimate the model using an exponential, a cubic, a linear spline, and a smooth quadratic spline functional form. Figure 7 shows the corresponding results.

We find that the key take-aways from the decomposition exercise are robust to functional form assumptions. First, gifted females are always referred less than gifted males. Second, the gender referral gap is mostly explained by males being more disruptive than females at high levels of IQs rather than due to differences in referral thresholds.

Figure 7: Validity: Different Functional Forms



Note: This figure plots the counterfactual exercise for different functional forms where $P_1(\hat{\alpha}_m, \hat{\tau}_m)$ and $P_1(\hat{\alpha}_f, \hat{\tau}_f)$ 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. $P_1(\hat{\alpha}_f, \hat{\tau}_m)$ shows the probability of referral for females, had they the same tolerance threshold as males. $P_1(\hat{\alpha}_m, \hat{\tau}_f)$ shows the probability of referral for females, had they the same IQ-disruption relationship as males.

5.2.2 Gender-specific IQ distributions

Our baseline analysis is based on the assumption that females and males have identical IQ distributions, both with mean 100 and standard deviation 15. While this assumption is backed up by the majority of the psychological literature, there is a strand of literature that challenges this assumption (Nyborg, 2015). Specifically, this literature brings forward the “greater variability hypothesis”, which states that the IQ distribution of males has a larger variance, which then would naturally increase the share of gifted males as compared to gifted females.

We address this concern in two ways. First, we re-estimate our model with the sex specific IQ distributions suggested by Nyborg (2015), that is $IQ|_{\theta_i=male} \sim N(103.19, 15.57^2)$ and $IQ|_{\theta_i=female} \sim N(102.18, 13.55^2)$. See Appendix ?? for resulting table of estimates and de-

composition figure. The estimation of this specification leads to a more pronounced quadratic relationship between IQ and disruption in the tails of the distribution. However, the main results do not change. We still find that the reason males are more likely to be referred than females is due to differences in the IQ-disruption relationship rather than differences in referral thresholds (see Appendix Figure 10). Note further, that assuming gender-specific IQ distributions worsens the model fit. Second, we turn around the question and assume that females and males have the same IQ-disruption relationship. We then ask, how different the IQ distributions would need to be to explain our empirical findings.¹⁸

To get at the gender-specific IQ distributions under the assumption of equal IQ-disruption relationships, we assume that $\alpha_m = \alpha_f$, and estimate the corresponding μ_m, μ_f, σ_m and σ_f to rationalize the observed gender-specific IQ-disruption patterns in the data. For details on this procedure, please see Appendix ???. The results of this exercise clearly show, that the assumption of equal IQ-disruption relationships is implausible. The exact replication of the above results would necessitate a negative variance of the IQ-distribution. Restricting the variance to be positive, the variance of the female IQ distribution would still need to be about 50 times larger than that of the male IQ distribution in order to fit the data. This extremely large difference is inconsistent with any evidence. To push this exercise further, we could instead restrict the variance of the IQ distributions to be the same, allow the quadratic term of the IQ-disruption relationship to be gender-specific, and back out how different the mean of the IQ distribution would have to be to rationalize the results. Doing this exercise, the mean of the female IQ distribution would be $\mu_f = 100$, while the one of the male distribution would be $\mu_m = 61.5$, which is again inconsistent with any evidence in the psychological literature.

This exercise nicely shows, that underlying differences in the IQ distribution between females and males cannot explain the observed gap in giftedness identification.

6 Conclusions

In most countries, the identification of intellectual giftedness is a two-step process. In a first step, a teacher refers the student to an assessor or school psychological service. In the second step, the student undergoes an IQ testing procedure and is subsequently diagnosed as being

¹⁸Note that the referral thresholds $\tau_{s,\theta}$ are identified using observed school-specific referral rates and observed disruption levels, and do therefore not rely on any assumptions on the IQ distribution. On the other hand, the estimated parameters α that define the IQ-disruption relationship do depend on the assumption of equal IQ distributions.

gifted, which is required before specialized education services can be delivered to the student. In this paper, we investigate two mechanisms that explain the underidentification of intellectual giftedness in female students within such a two-step identification process. 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 externalizing behaviors that will make them referred. This explains roughly 70% of the overall gender gap. On the other hand, 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, they confirm that two-stage identification systems for identifying intellectually gifted children are biased against female students. This has been suggested in the literature in the context of income and ethnicity (Card and Giuliano, 2016a,b; Peters and Matthews, 2016). A possible solution to increase screening fairness would require replacement or adjustment to the two-stage referral procedures for giftedness identification. One example is universal screening, which would not only be advantageous for female students, but could also reduce disadvantages for migrant and low-SES children. Universal screening has also been suggested in related but different settings, e.g. in screening for school-readiness (Balestra, Eugster, and Liebert, 2020). However, this solution is expensive for schools and educational systems.

Alternatively, schools and principals could standardize referral criteria to reduce subjectivity and especially adapt their referral guidelines to achieve gender-neutral referral probabilities. Referral should therefore rely on observable characteristics that are not gender-biased, especially not classroom disruption as we show in this paper differs systematically between male and female students. Specifically, it should be avoided to use subject-based markers (e.g. STEM vs verbal skills), behavioral stereotypes, or interests in specific subjects (e.g. robotics vs. reading). Instead, teachers might be trained to recognize consistently advanced cognitive performance (e.g. performance ahead of age-level expectations), precocious learning behaviors (e.g. the quick generalization of learnings to new contexts), advanced problem-solving skills and creativity, signs of asynchronous development across cognitive and socio-emotional domains, or discrepancy heuristics (e.g. high problem-solving ability but boredom-driven disengagement).

Finally, to alleviate the concern of under-identification of gifted female students, school curricula could be adapted in ways to serve all children better, no matter if they are intellectually

gifted or not. Examples for such adaptations would be the offering of enrichment lessons, where subjects can be studied in independent research projects with varying depth. Furthermore, more creative content such as creative writing workshops, robotics projects or design labs can be adapted to fit all abilities.

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.
- Bartels, Meike, Anne Hendriks, Matteo Mauri, Eva Krapohl, Alyce Whipp, Koen Bolhuis, Lucia Colodro Conde, Justin Lunningham, Hill Fung Ip, Fiona Hagenbeek et al. 2018. “Childhood aggression and the co-occurrence of behavioural and emotional problems: results across ages 3–16 years from multiple raters in six cohorts in the EU-ACTION project.” *European child & adolescent psychiatry* 27 (9):1105–1121.
- 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.
- Crijnen, Alfons AM, Thomas M Achenbach, and Frank C Verhulst. 1997. "Comparisons of problems reported by parents of children in 12 cultures: Total problems, externalizing, and internalizing." *Journal of the American Academy of Child & Adolescent Psychiatry* 36 (9):1269–1277.
- Deary, Ian J, Paul Irwing, Geoff Der, and Timothy C Bates. 2007. "Brother–sister differences in the g factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979." *Intelligence* 35 (5):451–456.
- Deary, Ian J, Graham Thorpe, Valerie Wilson, John M Starr, and Lawrence J Whalley. 2003. "Population sex differences in IQ at age 11: The Scottish mental survey 1932." *Intelligence* 31 (6):533–542.
- Derks, Eske M, James J Hudziak, and Dorret I Boomsma. 2007. "Why more boys than girls with ADHD receive treatment: a study of Dutch twins." *Twin Research and Human Genetics* 10 (5):765–770.
- 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.
- Exley, Christine L. and Judd B. Kessler. 2022. "The Gender Gap in Self-Promotion." *The Quarterly Journal of Economics* 137 (3):1345–1381.
- Giofrè, David, Enrico Toffalini, Lorenzo Esposito, and Cesare Cornoldi. 2024. "Sex/gender differences in general cognitive abilities: an investigation using the Leiter-3." *Cogn Process* 25 (4):663–672.
- Guénolé, Fabian, Jacqueline Louis, Christian Creveuil, Jean-Marc Baleyte, Claire Montlahuc, Pierre Fournier, 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." .
- Hinshaw, Stephen P, Phuc T Nguyen, Sinclair M O'Grady, and Emily A Rosenthal. 2022. "Annual Research Review: Attention-deficit/hyperactivity disorder in girls and women: underrepresentation, longitudinal processes, and key directions." *Journal of Child Psychology and Psychiatry* 63 (4):484–496.

- Hyde, Janet S, Sara M Lindberg, Marcia C Linn, Amy B Ellis, and Caroline C Williams. 2008. "Gender similarities characterize math performance." *Science* 321 (5888):494–495.
- Hyde, Janet Shibley. 2005. "The gender similarities hypothesis." *American Psychologist* 60 (6):581–592.
- Lavy, Victor and Yoav Goldstein. 2022. "Gifted Children Programs' Short and Long-Term Impact: Higher Education, Earnings, and the Knowledge-Economy." .
- Lavy, Victor and Analia Schlosser. 2011. "Mechanisms and impacts of gender peer effects at school." *American Economic Journal: Applied Economics* 3 (2):1–33.
- 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. 2025. "Mis (sed) diagnosis: Physician decision-making and ADHD." *Journal of Political Economy*, Conditional Acceptance.
- 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.
- Mowlem, Florence, Jessica Agnew-Blais, Eric Taylor, and Philip Asherson. 2019. "Do different factors influence whether girls versus boys meet ADHD diagnostic criteria? Sex differences among children with high ADHD symptoms." *Psychiatry research* 272:765–773.
- Neisser, Ulric, Gwyneth Boodoo, Thomas J Bouchard Jr, A Wade Boykin, Nathan Brody, Stephen J Ceci, Diane F Halpern, John C Loehlin, Robert Perloff, Robert J Sternberg et al. 1996. "Intelligence: knowns and unknowns." *American psychologist* 51 (2):77.
- 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." Tech. rep., National Bureau of Economic Research.
- 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.
- Strand, Steve, Ian J Deary, and Pauline Smith. 2006. “Sex differences in cognitive abilities test scores: A UK national picture.” *British Journal of Educational Psychology* 76 (3):463–480.
- 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.
- Vergunst, Francis, Melissa Commisso, Marie-Claude Geoffroy, Caroline Temcheff, Martine Poirier, Jungwee Park, Frank Vitaro, Richard Tremblay, Sylvana Côté, and Massimiliano Orri. 2023. “Association of childhood externalizing, internalizing, and comorbid symptoms with long-term economic and social outcomes.” *JAMA network open* 6 (1):e2249568–e2249568.
- Verhulst, Frank C, Thomas M Achenbach, Jan Van der Ende, Nese Erol, Michael C Lambert, Patrick WL Leung, Maria A Silva, Nelly Zilber, and Stephen R Zubrick. 2003. “Comparisons of problems reported by youths from seven countries.” *American Journal of Psychiatry* 160 (8):1479–1485.
- 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.

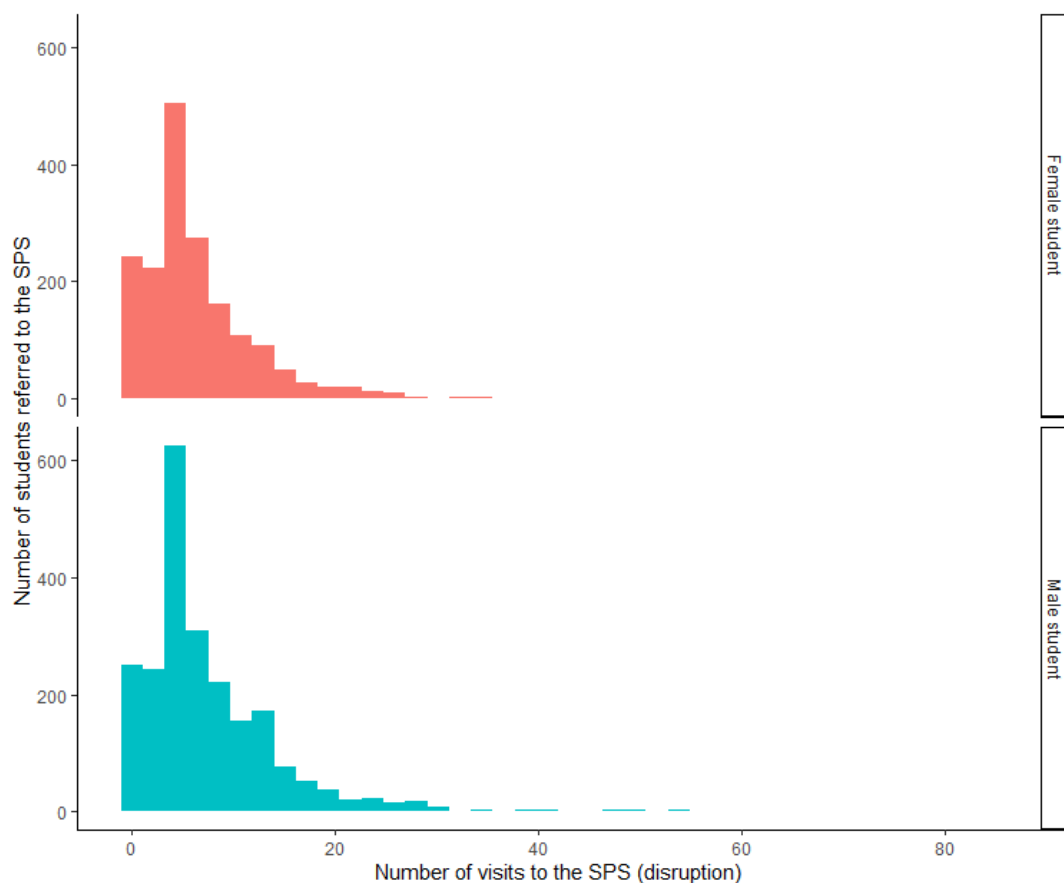
A Additional tables and figures

Table 3: 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

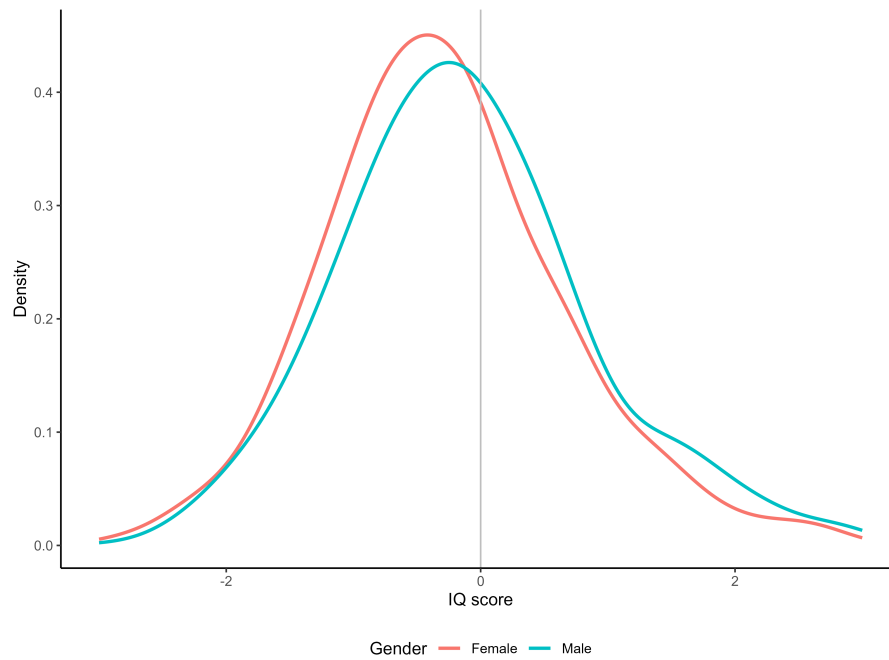
Notes: This table reports summary statistics about our reasons for referral in our sample. All students born between 1997 and 2000 and referred to the School Psychological Services (SPS) in the Swiss Canton of St. Gallen are included in the sample. 104 distinct school-cohort. *Source*: Ministry of Education,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

Figure 8: Distribution of number of visits to the SPS per gender



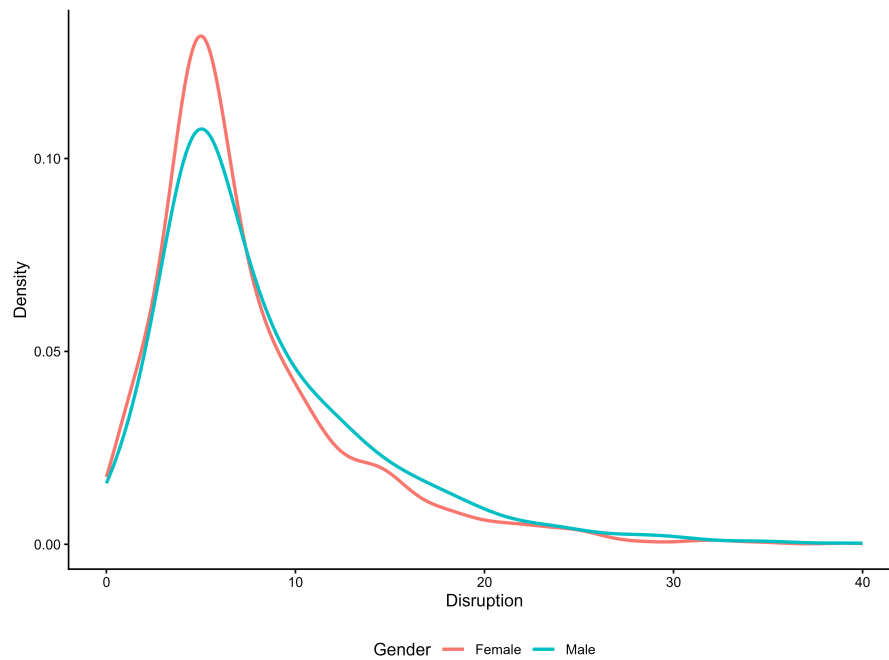
Note: This figure plots the distribution of the number of visits to the SPS per gender in the unrestricted sample (sample before restricting to cohort and schools of interest). *Source*: Ministry of Education, Canton of St. Gallen.

Figure 9: IQ distribution by gender



Note: This figure plots the IQ distribution by gender. *Source:* Ministry of Education, Canton of St. Gallen.

Figure 10: Distribution of disruption by gender



Note: This figure plots the distribution of disruption by gender. *Source:* Ministry of Education, Canton of St. Gallen.

B Gender and disruption

Throughout the paper, we use the number of visits to the caseworker as a proxy for disruption. In this section, we investigate whether this proxy might hide gender heterogeneity if, for instance, for the same number of visits to the caseworker, the intensity of the visits is higher for males than for females.

For each visit, we are able to observe the notes written by the psychologist. We leverage this information to capture the intensity of the visits, using both the type-token ratio and Yule’s I as richness measures, traditional measures in the natural language processing (NLP) literature. The type-token ratio (TTR) is a simple measure of lexical diversity in a text, calculated by dividing the number of distinct words (types) by the total number of words (tokens). Yule’s I quantifies how varied a text’s vocabulary is by inversely capturing its rate of repetition of words. We remove all stop-words and punctuation from the original raw text.

We find an average of 18.1 words per visit (13.6 words per visit after pre-processing/cleaning the text). Female students have on average 0.8 words less per visit than male students. Importantly, when comparing TTRs between female and male students and keeping the number of visits constant, we found no statistically significant differences in text richness between both groups. We obtain similar results using Yule’s I.

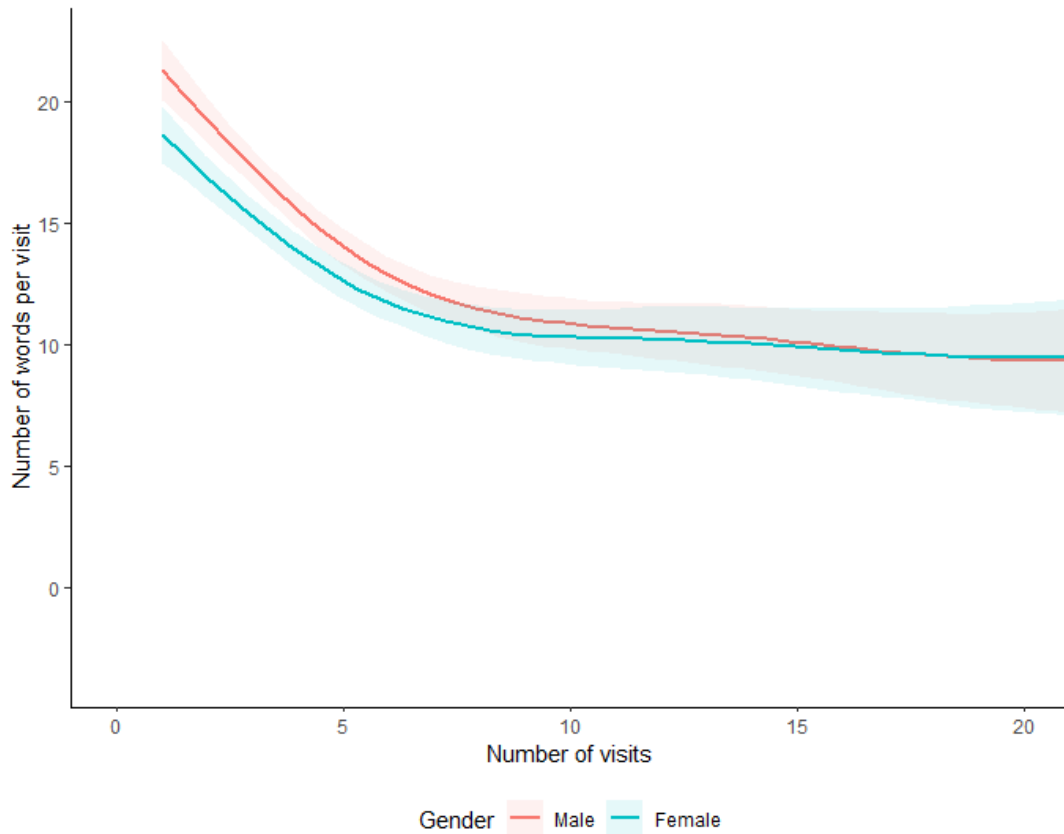
We also find a strong positive correlation between the number of words and the number of visits (an additional visit leads to 9.2 more words), and a small negative correlation between the number of words per visit and the number of visits (an additional visit leads to -0.4 more words per visit). This suggests that less text is written per visit for children with more difficult cases (more visits). Again, we find that these correlations are similar between male and female students. As can be seen in Figure 11, the number of words per visit decreases at the same rate for male and female students up to around 9 visits. For more difficult cases (above 9 visits), the difference in levels between male and female students disappears, plateauing at around 10 words per visit.

Table 4: Text differences across gender

	Average	Male Students	Female Students
Number of visits	7.16 (6.2)	7.7 (6.8)	6.5 (5.3)
Number of words per visit	18.1 (17.8)	18.7 (18.8)	17.4 (16.5)
Number of words (cleaned) per visit	13.6 (12.3)	14.0 (12.7)	13.1 (11.6)
Number of unique words per visit	15.0 (13.7)	15.3 (14.2)	14.5 (13.1)
Type-Token Ratio	0.89 (0.083)	0.89 (0.084)	0.90 (0.082)
Yule's I	166 (179)	171 (181)	160 (175)

Notes: This table reports summary statistics about text richness per visit. We use both the type-token ratio and Yule's I as richness measures. The type-token ratio (TTR) is a simple measure of lexical diversity in a text, calculated by dividing the number of distinct words (types) by the total number of words (tokens). Yule's I quantifies how varied a text's vocabulary is by inversely capturing its rate of repetition of words. We remove all stop-words and punctuation from the original raw text. *Source*: Ministry of Education, Canton of St. Gallen

Figure 11: Correlation: number of words per visit and number of visits per gender



Note: This figure plots the correlation between the number of words per visit and the number of visits per gender. *Source*: Ministry of Education, Canton of St. Gallen.