Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1120

The Correlated Proxy Problem: Why Control Variables can Obscure the Contribution of Selection Processes to Group-Level Inequality

Joao M. Souto-Maior
Stanford University

Kenneth A. Shores
University of Delaware

Rachel E. Fish
Smith College

Whether selection processes contribute to group-level disparities or merely reflect pre-existing inequalities is an important societal question. In the context of observational data, researchers, concerned about omitted-variable bias, assess selection-contributing inequality via a kitchen-sink approach, comparing selection outcomes of different-group individuals net of various characteristics. We introduce a conceptual framework that clearly defines the quantity of interest and argue that researchers should only control for the extent to which individuals meet selection criteria. Informed by this framework, we use directed acyclic graphs and structural equation modeling to show that traditional practices can inaccurately represent selection-contributing inequality because chosen controls frequently capture selection-irrelevant characteristics, which we define as the correlated proxy problem. Using Black-White disproportionality in special education as a case study, we show that typical practices of using test scores as covariates likely drastically underestimate the influence of selection-contributing inequality to Black over-representation in special education.

VERSION: January 2025

Suggested citation: Souto-Maior, Joao M., Kenneth A. Shores, and Rachel E. Fish. (2025). The Correlated Proxy Problem: Why Control Variables can Obscure the Contribution of Selection Processes to Group-Level Inequality. (EdWorkingPaper: 25 -1120). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/v3kz-g882

The Correlated Proxy Problem: Why Control Variables can Obscure the Contribution of Selection Processes to Group-Level Inequality

Joao M. Souto-Maior[1,4], Kenneth A. Shores[2], and Rachel E. Fish[3]

[1] Stanford University
[2] University of Delaware
[3] Smith College

[4] Corresponding author.
Email: joaosm@stanford.edu.
Address: 365 Lasuen St, Stanford, CA 94305.

December 17, 2024

Statements and declarations

Acknowledgments. We thank Ravi Shroff for helpful conversations.

Declaration of conflicting interest. The authors declare no conflict of interest.

Funding statement. The authors gratefully acknowledge from the WT Grant and Spencer Foundations for supporting this work.

Ethical approval and informed consent statements. Not applicable.

Data availability statement. The paper uses data from 1998 Early Childhood Longitudinal Study - Kindergarten cohort (1998 ECLS-K) kindergarten-eighth grade public-use file, publicly available for download at `https://nces.ed.gov/ecls/dataproducts.asp`. The code to fully reproduce all results in this paper is available from the corresponding author on reasonable request and will be made publicly available (in a GitHub repository) upon publication.

Abstract

Whether selection processes contribute to group-level disparities or merely reflect pre-existing inequalities is an important societal question. In the context of observational data, researchers, concerned about omitted-variable bias, assess selection-contributing inequality via a kitchen-sink approach, comparing selection outcomes of different-group individuals net of various characteristics. We introduce a conceptual framework that clearly defines the quantity of interest and argue that researchers should only control for the extent to which individuals meet selection criteria. Informed by this framework, we use directed acyclic graphs and structural equation modeling to show that traditional practices can inaccurately represent selection-contributing inequality because chosen controls frequently capture selection-irrelevant characteristics, which we define as the correlated proxy problem. Using Black-White disproportionality in special education as a case study, we show that typical practices of using test scores as covariates likely drastically underestimate the influence of selection-contributing inequality to Black over-representation in special education.

Keywords: selection processes; group-level inequalities; statistical controls; included-variable bias.

1 Introduction

Social stratification research is largely concerned with documenting and explaining differences in outcomes across individuals of different social categories (e.g., different outcomes for different-class, different-race or different-gender individuals), what are known as group-level inequalities (Grusky 1994; Hout and DiPrete 2006; Massey 2007; Shores et al. 2020; Tilly 1998). In this paper, we are concerned with the quantitative methodological problem of identifying whether a selection process is *contributing* to group-level inequality as opposed to simply *reflecting* group-level inequalities that have emerged, perhaps over a long history, prior to the selection process itself — what we refer to as *ex ante* inequalities (Goel et al. 2016). For the sake of terminology, we say that we are interested in understanding whether a selection process produces *selection-contributing* and/or *selection-reflecting* inequalities across different-group individuals.[1]

[1] Some of the existing literature, such as assessments of hiring processes (Quillian and Midtbøen 2021), rely on the concept of *discrimination*. Our view is that this concept might be too narrow to capture the quantity of interest. Discrimination is often associated with the actions of gatekeepers; i.e., is usually understood that discrimination exists only if it can be shown that gatekeepers give different decisions for different-category individuals which are similar across relevant characteristics (Lucas 2009). However, selection processes are multi-staged and the decisions of gatekeepers is only one of such stages. Hiring processes, for example,

Take, as a simple example, an exam for a driver's license. Two students from different schools go to take the exam. Student A (male) attends an all male school where driver's education is emphasized, whereas student B (female) attends an all female school where driver's education is not. Consequently, at the time of the exam, student A has better driving abilities than student B. Suppose that student A passes the exam and student B fails. In this case, we observe gender disparities in exam-passage outcomes, but the selection mechanism simply *reflects* inequalities in opportunities for preparation (and, hence, only produces a selection-reflecting inequality).

In contrast, suppose that students A and B had equal preparation and were actually equally qualified to pass the exam. Suppose, further, there is a human grader for the driving portion of the exam that systematically evaluates female drivers more stringently, leading student A to pass and student B to fail. In this case, we again observe gender disparities in exam-passage outcomes but the selection mechanism *contributes* to inequality (and, hence, produces a selection-contributing inequality).

Understanding whether selection processes contribute to (versus merely reflect) inequality has important policy implications, though scrutiny of these policy implications is warranted. If a selection process simply reflects pre-existing inequalities, then equity-focused interventions should prioritize addressing disparities that emerge before selection decisions. When selection processes actively contribute to inequality, reforming these decision mechanisms becomes an additional avenue for intervention. We note here that this focus on selection processes, while politically expedient because they appear more readily changeable, may divert attention from deeper structural inequities that ultimately generate larger disparities (e.g., Royce 2022). Beyond immediate policy implications, selection-contributing inequality remains a significant marker of inequity, as discriminatory gatekeeping provokes particular

involve the processes via which individuals (a) learn about job openings; (b) chose to apply for a position; (c) are selected for a call-back; and, finally, (d) receive a job offer. If we are interested in identifying discrimination per se in the hiring process, then analyses are restricted to stages (c) and (d). Here, we avoid the concept of discrimination and, rather, rely on the notion of *selection-contributing inequality* to dialogue with quantitative assessments of group favoritism across any (or all) stages of selection processes.

moral concern since it violates basic principles of procedural fairness and equal treatment (e.g., for a review Lind and Tyler 2013; MacCoun 2005; Miller 2001). For instance, people often view unfair selection mechanisms as especially objectionable because they represent active choices to perpetuate inequality rather than passive reflection of existing disparities (Akbaş et al. 2019).

Informed by these policy motivations, social scientists have long been interested in quantitatively identifying whether (and to what extent) various selection processes contribute to group-level inequalities. Examples of such selection processes include, but are not limited to, hiring processes (Brand 2015; Pedulla 2020; Quillian and Midtbøen 2021); within-firm promotions (Rivera and Tilcsik 2019; Skaggs and Bridges 2013); mortgage-lending decisions (Pager and Shepherd 2008; Quillian et al. 2020); police-stopping decisions (Goel et al. 2016; Legewie 2016; Ray et al. 2024); child protected services determinations (Feely and Bosk 2021); health care provision and treatment decisions (Spencer and Grace 2016), college admissions decisions (Jayakumar and Page 2021; Warikoo 2022); disciplinary decisions within schools (Muñiz 2021; Skiba et al. 2016b); and selection into educational services/programs (Domina et al. 2017) such as advanced courses (Lucas et al. 2020; Mickelson 2001), gifted and talented programs (Pearman and McGee 2022; Peters et al. 2019) and special education services (Fish 2017; Harry and Klingner 2014; Mayes 2023).

Quantifying selection-contributing inequality is a challenging task. Due to historical and structural sources of group-level inequality (Massey 2007; Tilly 1998), groups might differ in substantial ways at the time in which selection processes take place, and, given a selection mechanism that values factors associated with group-level differences, these differences might explain differential selection outcomes. Thus, to identify selection-contributing inequality the researcher must, ideally, compare the selection outcomes of different-group individuals who are similar across *relevant* ex ante characteristics. This type of comparison presents two complications: first, the researcher must clearly define which ex ante differences are relevant for the comparison of interest; second, the researcher needs to quantitatively operationalize such relevant characteristics, many of which are unlikely to be empirically observable.

Methodologically, field experiments are known as the best available strategies to tackle these issues (Pager and Shepherd 2008; Quillian and Midtbøen 2021) and have provided a significant contribution to social scientific understanding of selection processes — see, for instance, audit studies of discrimination in hiring processes (Gaddis 2018). However, due to practical and ethical concerns, the possibility of field experiments across the many domains of interest is limited (Pager and Shepherd 2008). Therefore, many (if not most) quantitative examinations of the extent to which selection processes contribute to inequality rely on observational data, leveraging available covariates to try and control for ex ante characteristics.

Among the fields mentioned above, consider, for instance, the following examples, which use covariates, such as test scores and/or socio-demographic variables, to try and identify group inequality due to selection mechanisms: hiring processes (Farber and Valletta 2015; Fernandez and Sosa 2005; Petersen et al. 2005; Petersen and Togstad 2006); within-firm promotions (Blau and DeVaro 2007; DiPrete and Soule 1988; Petersen and Saporta 2004) mortgage-lending decisions (Gotham 1998); police-stopping decisions (Gelman et al. 2007); health care provision and treatment decisions (Baglivio et al. 2017; Balsa et al. 2005), child protected services determinations (Font et al. 2012; Putnam-Hornstein et al. 2013; Rivaux et al. 2008) college admissions decisions (Arcidiacono et al. 2022; Espenshade et al. 2004; Grossman et al. 2024b); disciplinary decisions within schools (Barrett et al. 2021; Edwards 2016; Huang 2020); and selection into educational services/programs such as advanced courses (Conger et al. 2009; Irizarry 2021; Kelly 2009; Riegle-Crumb et al. 2019), gifted and talented programs (Grissom and Redding 2015; Long et al. 2023) and special education services (Fish 2019; Hibel et al. 2010; Morgan et al. 2015; Skiba et al. 2005).

As evidenced by these various studies, the traditional approach of leveraging observational data to assess selection-contributing inequality is to estimate a regression model that compares the selection outcomes of individuals from different groups that are similar on a set of relevant characteristics. However, the notion of relevant characteristics is rarely clearly defined. Most commonly, scholars essentially adopt a *kitchen-sink* approach; that is, they statistically account for a series of available measures that differ across groups that are also associated with outcomes. One can understand why this approach is appealing. If the researcher fails to control for relevant group-level differences that matter for outcomes related to the selection mechanism, then estimates will fail to distinguish between the ex ante causes of the inequality and the selection mechanism itself. Failing to make this distinction is a form of omitted-variable bias — i.e., absent controls, different social category individuals have different exposures to structural inequalities, causing unequal ex ante probabilities of meeting this selection mechanism's requirements, resulting in different outcomes unrelated to the selection mechanism itself. Controlling for various measures minimizes this risk. It is this traditional regression-based approach of using observational data to assess selection-contributing inequality that our paper targets.

We argue and illustrate that while this kitchen-sink approach has dominated observational studies around this question, it is vulnerable to two important, but often unrecognized, methodological problems. First, like many quantitative investigations (Lundberg et al. 2021), it lacks a conceptual framework to explicitly define the statistical quantity of interest (the estimand) and to define what the relevant characteristics of interest are. Second, perhaps due to the lack of such a conceptual framework, concerns for omitted-variable bias have overlooked another, but less-often-recognized, potential issue: the fact that the *inclusion* of control variables might also bias estimates if different-group individuals who appear dissimilar net of control variables are, in fact, similar across all relevant characteristics, what we call the correlated proxy problem. While noted by selected statistical studies (Ayres 2005, 2010; Jung et al. 2024), this issue is yet to be formally and intuitively described. We call this issue *the correlated proxy problem.*[2]

[2]The methodological issue we describe here has been originally introduced by statisticians under the terminology of "included-variable bias" (Ayres 2010). However, we choose to refer to it as *the correlated proxy problem* because the umbrella term of included-variable bias can be generically used to represent a host of instances of "bad controls", such as the problem of controlling for mediators (Sampson 2008) and of

Informed by these issues, the goals of this paper are (1) to provide a conceptual framework that clearly defines the estimand of interest in quantitative assessments of selection-contributing inequality; and, informed by this conceptual framework, (2) to formally (and intuitively) describe how this often-neglected statistical issue might arise.

To make our analyses more concrete, the paper is structured around a running research question: to what extent is Black-White inequality in special education receipt attributable to the selection process by which students are identified to receive special education services (selection-contributing inequality) versus Black-White differences in ex ante selection-relevant characteristics (selection-reflecting inequality)? We choose this case study for two reasons. First, this question has clear policy implications, and has been a source of lively academic and policy debate (Cruz and Rodl 2018; Morgan and Farkas 2015; Morgan et al. 2015; Skiba et al. 2016a). Second, federal guidelines, described below, require educators to make precisely the distinction that researchers have struggled to capture: determining whether a student's characteristics, such as lower academic performance, stem from unequal educational opportunity (a selection-irrelevant cause) or from a disability (a selection-relevant cause). This alignment between federal requirements and our theoretical framework allows us to use special education as an illustrative case for examining how researchers might better approach the measurement of selection-contributing inequality.

In what follows, we first situate our study by linking it to related methodological literature and summarizing our running example of disproportionality in special education. We then provide a conceptual framework that clearly defines the quantity of interest in assessments of selection-contributing inequality. We formalize this conceptual framework and define the bias using linear simultaneous equation models (LSEMs) and directed acyclic graphs (DAGs). Finally, we illustrate and assess the correlated proxy problem using the example of racial disproportionality in special education.

---

controlling for post-treatment effects (Rosenbaum 1984).

2 Background

Relationship to Literature on Control Variable Selection

Two strands of methodological research inform our analysis of selection-contributing inequality. First, the "bad controls" literature demonstrates how controlling for certain variables can introduce bias in causal estimates, even absent traditional omitted variable concerns (Cinelli et al. 2024; Elwert and Winship 2014). This research identifies several mechanisms through which controls bias estimates: mediation bias when controls lie on the causal path between treatment and outcome (Sampson 2008), post-treatment bias from controlling for variables affected by treatment (Rosenbaum 1984), collider bias when conditioning on common effects (Pearl 2000), and various forms of M-bias arising from specific causal structures (Ding and Miratrix 2015). The back-door criterion synthesizes these insights by formalizing conditions under which controls yield unbiased causal estimates: controls should block non-causal paths while preserving causal paths between treatment and outcome (Cinelli et al. 2024).

Our framework shares this literature's goal of isolating relationships of interest through careful control variable selection. However, our focus is on a different estimand —selection-contributing inequality— which implies measuring how selection mechanisms contribute to group differences in outcome Y. As we will argue in this paper, this estimand requires researchers using observational data to control for between-group differences in selection-relevant characteristics (Z), such as the ability to drive effectively in our earlier example. The bad controls literature, however, would typically flag Z as a problematic control since it mediates the relationship between group membership and outcomes. Thus, while the "bad controls" literature has provided tools for thinking rigorously about control variable selection, assessing selection-contributing inequality requires an adaptation of this framework.

The mediating role of selection-relevant characteristics connects our work to a second strand of literature: causal mediation analysis. This approach, developed by Baron and Kenny (1986) and extended by others (e.g., Imai et al. 2010; Pearl 2014), decomposes total causal effects into direct and indirect pathways. Traditional mediation analysis separates the direct effect ($T \to Y$) from the indirect effect ($T \to M \to Y$). Similarly, our framework decomposes group disparities into selection-contributing inequality ($G \to Y$, not through Z) and selection-reflecting inequality ($G \to Z \to Y$). This decomposition allows us to isolate disparities arising from selection mechanisms versus those reflecting pre-existing qualification differences.

Here, as well, our setting presents a key distinction, as the latent variable Z (e.g., representing the unobserved ability to effectively drive a car) is inherently unobservable. Thus, researchers face an unavoidable trade-off: they must include proxies to address confounding while avoiding ones that conflate selection-contributing and selection-reflecting inequalities. Using methods from both literatures—directed acyclic graphs (DAGs) and linear structural equation modeling (LSEM)—we demonstrate how attempts to proxy for Z can confound these two components of between-group inequality, even when following best practices for control variable selection.

Running example: racial disproportionality in special education

Descriptive analyses consistently show that racially minoritized students, and Black students especially, receive special education services, measured as the receipt of an Individual Educational Program (IEP), at higher rates than White students, with the majority of research focusing on Black-White inequality (Artiles 2013; Donovan and Cross 2002; Dunn 1968). These disparities have been of concern to researchers and policymakers for decades, and federal policy has required states to monitor and address this "disproportionality" since 2004 (IDEA-Sec.300.309 2006). The prevailing wisdom has long been that the over-representation of Black students (relative to Whites) is attributed to two main processes: racial bias in disability identification via school-based processes (Artiles et al. 2002; Donovan and Cross 2002; Fish 2017; Harry and Klingner 2014; Mayes 2023)— what is, to use the terminology proposed here, a selection-contributing inequality — and structural inequalities that lead Black students to be, on average, more likely than White students to have disabilities qualifying them for special education services (Artiles et al. 2002; Donovan and Cross 2002; Shifrer 2018) — a selection-reflecting inequality.

Just as in the observational methods described above, quantitative researchers try to distinguish between these competing explanations by estimating a regression model of students' IEP receipt status on race and a series of covariates, such as socioeconomic status and levels of academic preparedness, usually encoded as test scores (e.g., Fish 2019; Hibel et al. 2010; Kincaid and Sullivan 2017; Morgan et al. 2012, 2015, 2020; Shifrer 2018; Sullivan and Bal 2013). Intuitively, this approach aims to contrast IEP receipt rates for Black and White students who are comparable in terms of economic disadvantage and achievement. Though the purposes of these covariates are not stated in the terminology we propose here (i.e., proxy variables), they can broadly be said to represent proxies for either disability itself (Hibel et al. 2010; Morgan 2021) or clinical need, implying that test scores and socio-demographic information are useful for understanding who should receive special education services (Moll et al. 2014; Shifrer 2018).

The conclusion drawn from this framework is that Black students are actually less likely to receive special education than comparable White students, as the conditional rate of IEP receipt for Black students is less than it is for White students given these covariates (see Morgan et al. (2017) for a synthesis of this literature). This result points to selection mechanisms that contribute to racial inequity in the form of *under*-representation of Black students in special education. These results have had a notable effect on policy, shifting focus away from reducing the unconditional over-representation of Black students in special education to identifying more Black students with disabilities to receive special education services (e.g., Barnum 2018; Burke 2024: p. 334).

We can contrast what researchers traditionally do to identify whether selection processes contribute to racial disproportionality in special education with what is legally required by federal special education policy under the *Individuals with Disabilities Education Act of*

2004 (IDEA 2004). To make our running example more concrete, consider the case of special education services for specific learning disabilities (SLDs), the most common disability category (U.S. Department of Education 2019). Under IDEA regulations, a student has a specific learning disability if the student does not meet the approved grade-level standards in at least one of 8 domains representing oral, listening, written, reading and mathematics abilities, and, notably, if such low academic performance is not explained by "exclusionary factors" — factors (such as inappropriate instruction, other disabilities, emotional disturbances, limited English proficiency, and cultural or economic disadvantages) that can create unequal educational opportunities (IDEA-Sec.300.309 2006). Essentially, IDEA states that if a student's below grade level academic performance is due to some other disability or unequal educational opportunity, then it is not a specific learning disability. Educators are then tasked with ruling out these other causes of disability before a designation is made.

Because of this academic and policy relevance, we believe this area of inquiry provides a valuable context for an empirical illustration of what we call the correlated proxy problem.

3 Conceptual framework

The quantity of interest: selection-contributing inequality

For generality, let us first introduce some terminology. There are a wide range of selection processes that can contribute to inequality. As discussed in the introduction, these selection processes can range from hiring decisions to college admissions. To capture these wide range of processes, let the language of "program" capture the resource, policy, service, or program for which individuals are being selected. Further, let *program disproportionality* refer to the under- or over-representation of some group A (relative to another group B) in a given program.

Recall that to identify whether a selection process itself contributes to group-level inequality (i.e., is *selection-contributing*), researchers would like to compare the selection outcomes of different-group individuals among those individuals with the same set of relevant characteristics. In our introductory example, we implicitly indicated that such relevant characteristics capture factors such as ability to drive (e.g., knowledge of rules and demonstration of parallel parking), and in our specific learning disability example, we used IDEA's definition to say that relevant characteristics include below academic grade performance not attributable to educational opportunities or other disabilities. Now, we build on this intuition to provide a general framework that clearly defines which characteristics are relevant in the comparisons of interest. To do so, it is necessary to clearly define our estimand, that is, what does it mean to say that a selection process contributes to (as opposed to simply reflects) inequality?

We illustrate this question through an example. Suppose an administrator observes that, in a given school, Black students are more likely to receive special education services (an IEP) for an SLD than White students and asks: are we, as school-level personnel, contributing to this inequality by, for example, over-estimating the incidences of SLDs for Black students? Or, in contrast, is it that structural differences in the life experiences of Black and White students (such as differences in exposure to lead and access to appropriate health care) have made it so that these groups of students currently have different incidences of learning disabilities? Essentially, what the educator wants to know is whether the school-implemented special education selection process is creating additional Black-White inequality in IEP receipt rates above the structural inequality that gave rise to differential incidences of special education needs in the first place.

Recall that, according to IDEA, we can define a learning disability as an academic difficulty that cannot be explained by exclusionary factors — i.e., factors which can negatively affect student learning but are considered distinct from other disabilities or lack of educational opportunities. Then, in this context, Black students with lower academic performance are not by default considered to have a learning disability; rather, school personnel must evaluate whether students' reduced academic performance is not caused by economic circumstances or/and lack of educational opportunities. Differentiating these mechanisms is the job of the referral team, and what our hypothetical administrator is asking is whether the referral team is effectively doing their job of following IDEA regulations to ensure exclusionary factors are not included as part of the referral process. To assess the administrator's question, researchers should, therefore, implement similar safeguards and also distinguish between these mechanisms.

This example illustrates a key principle: to determine whether a selection process contributes to (rather than merely reflects) group-level disparities in program access, we must compare individuals that similarly meet the criteria that should govern selection.[3] In our case, this means comparing students with the same severity of learning disability. Group-level differences in program access that persist even when comparing individuals with equivalent levels of the relevant underlying characteristic would indicate that the selection process itself, rather than pre-existing differences in qualification, is contributing to observed disparities. Therefore, to properly assess the selection-contributing effect, we must identify and control for these criterion-relevant characteristics—those that capture legitimate differences in qualification for the program.

To operationalize this principle, researchers should therefore follow two key steps: (a) quantify the latent concept that captures the extent to which different-group individuals meet the selection criteria, and (b) compare selection outcomes between groups at equivalent levels of this latent measure. If group-level disparities in program access are fully explained by differences in the latent measure, this indicates selection-*reflecting* inequality. Conversely, disparities that persist after accounting for these criteria-related differences provide evidence of selection-*contributing* inequality.

To illustrate this conceptual framework, let us continue with the questions raised by the administrator. Suppose that some objective measure of learning disability could be created and that the administrator could apply a universal diagnostic tool to compute this measure for all students. Then, the administrator could observe whether Black students

[3] This framework borrows insights from the legal perspective of discrimination known as disparate impact (for an overview, see Lucas 2009). See Goel et al. (2016) for this argument in the context of racial discrimination in stop-and-frisk practices based on an statistical interpretation of legal requirements.

with (or without) a learning disability are (or are not) receiving special education services; a similar exercise could be conducted for White students. Table 1 presents a hypothetical data resulting from such an exercise.

Table 1: Hypothetical data for the assessment of selection-contributing inequality in the receipt of special education services for specific learning disability (SLD).

	With a learning disability	Without a learning disability
***Panel A**: Sample Statistics*		
Black Students		
With an IEP	20	4
Without an IEP	1	40
White Students		
With an IEP	40	14
Without an IEP	2	139
***Panel B**: IEP and Disability Rates*		
Black Students		
Proportion with a learning disability		32.3% (21/65)
Proportion receiving an IEP		36.9% (24/65)
Proportion under-served		4.8% (1/21)
Proportion over-served		9.1% (4/44)
White Students		
Proportion with a learning disability		21.5% (42/195)
Proportion receiving an IEP		36.9% (54/195)
Proportion under-served		4.8% (2/42)
Proportion over-served		9.1% (4/44)

In this stylized example, raw disproportionality in IEP receipt rates is about 10 percentage points, but because true learning disability rates are observed, it can be seen that neither Black nor White students are different in their rates of over- or under-receiving services; rather, the difference in IEP receipt rates is attributable to the difference incidences of a learning disability between groups. Using language introduced above, we would say that, in the context of these hypothetical data, the administrator should conclude that the raw racial disproportionality reflects a selection-reflecting inequality, i.e., gaps in IEP receipt rates are fully explained by ex ante criteria-related differences.

Two fundamental measurement problems

This conceptual framework helps clarify that addressing this research question implies two measurement problems, which are central to methodological issue we are concerned with in this paper.

The first measurement problem is that the determination of the selection criteria is *normative*. Here we mean that other selection criteria might be suggested that could reveal different levels of selection-contributing inequality, holding all the same data fixed. For example, in the segregated southern part of the United States prior to Brown, race would have been a selection criteria to determine program eligibility. Then, clearly, comparing program participation rates among those meeting the selection criteria — in this case, race itself —- would reveal no selection-contributing inequality across racial lines. It is clear from this example that some selection-relevant criteria can be normatively wrong. Other, more contested, examples exist. For enrollment into advanced placement (AP) courses[4], a criteria condition might be academic preparedness — a probability of succeeding in an AP class — or, alternatively, having taken a set of prerequisite courses, such as pre-calculus for AP calculus. These different criteria conditions are disputable and could provide different accounts of selection-contributing inequality if utilized (Souto-Maior and Shroff 2024).

In the empirical literature on selection-contributing inequality, researchers typically employ control variables without explicitly stating what latent selection criteria these controls are meant to represent, or, even, whether the implied criteria are legitimate requisites of program access. Studies generally move directly to statistical estimation without first establishing what constitutes a valid basis for group differences. Our theoretical framework demonstrates why this omission matters: without clearly specifying the assumed selection criteria, researchers cannot properly define their estimand of interest or appropriately measure the extent to which selection processes contribute to group-level inequality.

[4] A set of courses which provide advanced, college-preparatory, curriculum to selected high school students in the US.

The second fundamental measurement problem is that, supposing a selection criteria could be agreed upon or at least proffered by the researcher or policymaker, it might not in practice be possible to measure it. In the case where the selection criteria include prerequisite courses, then the criteria is observable with transcript data, and the researcher can condition on students with matching prerequisites. In other cases — perhaps the modal case — the selection criteria is latent or not directly observable. In our running example, we acted as if the administrator could go out and evaluate all students to determine whether some learning disability existed or not. However, no such diagnostic exists, for many reasons, perhaps most relevantly is that disability cannot be directly observed (Fletcher et al. 2018), and the definitions of disability are themselves contested (Maroto et al. 2019; Shakespeare 2013; Shifrer and Frederick 2019).

Given the absence of directly observable selection criteria, researchers often rely on proxies for one (though as we note, the specific latent variable for which proxies are used is rarely explicated). For special education research, as we described before, test scores and socio-demographic variables might be used as proxies for disability itself or as evidence of clinical need for special education services. Further, given the absence of a direct measure of the selection-relevant criteria and well-known concerns about omitted-variable bias (Angrist and Pischke 2009; Farkas and Vicknair 1996; Morgan 2021), it is common for researchers to saturate their regressions with as many available covariates as possible. However, this conventional approach may capture group-level differences unrelated to the selection criteria itself, risking what we call the correlated proxy problem. Below, we formalize what this risk is.

4 Formalization of our conceptual framework

We now formalize this conceptual framework to clearly articulate and illustrate what we mean by the correlated proxy problem.

Suppose that, for a population of N individuals, the researcher is interested in identifying the extent to which the selection mechanism for a given program Y (e.g., special education services) produces selection-contributing inequalities with respect to group membership G (e.g, race). Following the framework proposed above, the researcher should account (statistically control) for the extent to which different-group individuals meet the selection criteria. For each student i, we have the following variables (subscript "i" omitted throughout to simplify notation):

- Y is a binary indicator for whether the student has access to the program.
- G is an indicator of i's group membership. For simplicity, we assume only two group categories: $G = 1$ if the individual is a member of the group of interest and $G = 0$ if the individual is a member of the reference group category. For example, suppose that G represents race, and that we are only interested in the "Black" and "White" racial categories. Using "White" as the reference category, we set $G = 1$ if i's race is "Black" and 0 if "White".

To illustrate in an empirical context, let us go back to our running example. In this context, Y captures a binary indicator for whether the student receives special education services, defined as an Individual Education Plan (IEP); G is a binary indicator of i's racial category where $G = 1$ if i's race is "White" and $G = 1$ if i's race is "Black".

The raw model

The simplest kind of evidence with respect to program disproportionality involves estimates of raw (or unconditional) disparities. These estimates capture the extent to which access to a given program varies across groups without accounting for any pre-existing group-level differences. These raw measures provide useful information. For instance, under the well-supported premise of no innate or biological race differences in incidences of disability (Mayes 2023), evidence of unconditional racial disproportionality in receipt of special education services points to the existence of social processes — inside and outside of school — that shape racial disparities in special education services (Donovan and Cross 2002; Fish 2017).

Raw group-level inequality in program Y with respect to group membership G can be computed by a regression of Y on G, what we call the *raw model*:

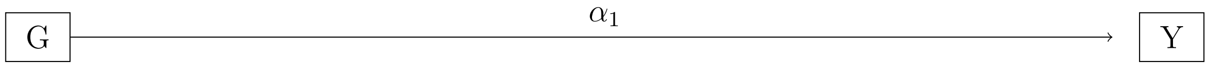
$$Y = \beta_1^{raw} G + \epsilon. \qquad \text{(raw model)}$$

Here we must make a small methodological aside. This regression model, as well as the subsequent models presented here, are written as linear probability models and so, are estimated through ordinary least squares (OLS). The coefficient on G can be interpreted as the average marginal difference between the category of interest (as opposed to the reference category) in the predicted probability of program enrollment. Many existing studies estimate these models through a logistic regression rather then a linear probability model. We choose a linear probability model (and the marginal difference/effects interpretation) because a central objective of this paper is to assess how the coefficient on G varies across the inclusion of different control variables. Unlike linear regression coefficients, logistic regression coefficients are affected by the level of unobserved heterogeneity in the model — they are influenced by omitted variables, "even when those variables are unrelated to the independent variables in the model" (Mood 2010: p. 67). Thus, with logistic regression, adding control variables changes how the model accounts for unexplained variation, making it impossible to attribute changes in coefficients solely to the substantive effect of those controls (Mize 2019; Mood 2010).

In the raw model, if $\hat{\beta}_1^{raw}$ is statistically different from 0, then we observe raw group-level inequality. This is the coefficient which earlier special education research focused on (Artiles et al. 2002; Donovan and Cross 2002; Harry and Klingner 2014; Skiba et al. 2005) and which consistently indicates a higher rate of special education receipt for Black students relative to White students.

From a DAG representation, this model estimates the pathway displayed in Figure 1, below:

Figure 1: A DAG representation of the pathway of interest when estimating raw group-level program disproportionality.



The ideal model

While informative, the raw model does not indicate whether students with the same selection-relevant characteristics are over- or under-represented in the outcome Y. Suppose a given variable (or set of variables), Z, captures the extent to which individual i meets the selection criteria. In the context of our running example, Z captures the extent to which student i has a learning disability (as defined by IDEA) prior to special education identification decisions. Without loss of generalizability and to align with latent variable measurement, we propose Z to be interpreted as the standardized probability (or risk) that student i has a learning disability.[5]

In an ideal, hypothetical, scenario in which Z is observable to the researcher, the data-generating process of interest can be described by the DAG in Figure 2, below:

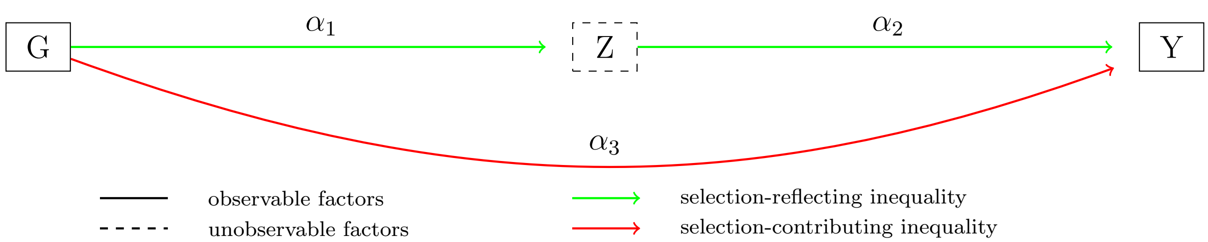


Figure 2: The data-generating process of interest if Z is observable by the researcher.

This DAG presented in Figure 2 illustrates the paths of interest through which differences in the outcome Y between group membership G can occur. First, there are pathways (α_1 and α_2) that pass through Z. Path α_1 captures potential group-level differences with respect to criteria-related characteristics and path α_2 represents the extent to which criteria-related

[5]While disability is typically coded as a binary category for administrative purposes, this reflects practical considerations rather than theoretical ones. We treat Z as continuous to avoid imposing arbitrary thresholds to dichotomize disability status.

differences influence selection outcomes. Using the terminology introduced earlier, this path only captures the source of selection-reflecting inequality, i.e., group-level inequalities that arise through this path result from group-level differences in how individuals meet the selection criteria or group-level differences in the relationship between the selection criteria and the outcome.

The remainder of group-level differences in program enrollment are captured by path α_3. This path captures group-level differences in Y that *cannot* be attributed to differences in how different-group individuals meet the selection criteria and, thus, following earlier terminology, captures the source of selection-contributing inequality. Therefore, α_3 is what we would like to estimate.

Returning to our running example, due to ex ante inequalities, such as exposure to environmental toxins and family stress (Cichy et al. 2012; Sampson and Winter 2016), it is possible that, at the time of school-based special education identification processes, Black students are, on average, more likely than White students to have a learning disability (path α_1). In addition, differences in learning disabilities might shape how gatekeepers make selection decisions, leading to group-level differences in special education receipt rates (path α_2), a further source of selection-reflecting inequality. However, if group-level differences in special education receipt are not based on differences in Z — that is, if group-level differences occur for students with similar Z — then path α_3 will arise, representing a selection-contributing inequality.

Group-level differences in incidences of learning disability might not fully explain disparities in special education receipt rates ($\alpha_3 \neq 0$) for numerous reasons; here, we highlight two. First, the mechanisms by which students are selected for special education programs are often subjective, giving strong decision power to gatekeepers — e.g., school-level personnel such as teachers, administrators and social workers — and thus, might be subject to pressures from families or biases from gatekeepers. Second, the actual selection mechanism used by schools when making special education decisions might not appropriately select on the basis of learning disability. For instance, most schools rely on at least some measure of student performance when making special education decisions, which may associate racial differences in educational opportunities along with disability rates.

In the ideal scenario where Z is observable, we can identify selection-contributing inequality by simply comparing the program enrollment rates of different-group individuals which are similar with respect to Z. The following regression captures this ideal strategy:

$$Y = \beta_1^{ideal} G + \beta_2 Z + \epsilon. \tag{ideal model}$$

If $\hat{\beta}_1^{ideal}$ is statistically different from 0, the empirical model would indicate the existence of selection-contributing inequality.

The proxy-based model

The challenge for the researcher is that, in practice, Z is (almost always) unknown. This means that we can only observe raw between-group inequality (the effect of G on Y as defined by the raw model), but we cannot estimate the ideal model to assess the extent to which it can be attributed to selection-reflecting and/or selection-contributing inequality.

Researchers are well-aware that it is problematic to infer that an $\hat{\beta}_1^{raw}$ which is statistically different from 0 implies a selection-contributing inequality. For this conclusion to be true, an unlikely scenario would need to hold. It would have to be the case that α_1 or α_2 in Figure 2's DAG were 0, allowing us to say that the selection criteria plays no role in explaining group-level differences in program enrollment. In this very specific case, $\hat{\beta}_1^{raw}$ would, in fact, only indicate selection-contributing inequality.

Given the structural foundations of group-level inequalities (Massey 2007; Tilly 1998), it is reasonable to assume that, in most contexts (such as our example of racial disproportionality in special education), individuals from different groups will likely have different values of Z at the time of selection decisions ($\alpha_1 \neq 0$) or that selection decisions are likely to be influenced by Z ($\alpha_2 \neq 0$).

Given that Z is unobservable, researchers often use a variable (or set of variables) as a proxy for Z. This is implicitly what the traditional kitchen-sink approach attempts to do. Let X represent the chosen control variable(s). Here, we acknowledge that covariate-selection practices in quantitative research provide some guidance for the selection of X (e.g., Cinelli et al. 2024). Specifically, X should not introduce problems traditionally associated with "bad controls", such as covariates containing group-level measurement errors or post-treatment effects (i.e., controls observed concurrently or after the selection process that are correlated with the regressor of interest (Gaebler et al. 2022)). Other examples from this literature, such as the problem of conditioning on a mediator, will not help researchers with the selection of X in our context, as the aim here is to isolate disparities that exist after accounting for selection-relevant criteria, which requires controlling for certain mediating pathways (Z). Thus, here and throughout, we reiterate that the data generating processes we characterize do not have the problems of group-level measurement errors or post-treatment controls. While X can be interpreted as a set of variables, for clarity our examples will focus on the case where X captures a single variable.

Then, the general practice is to assess selection-contributing inequality via what we call the *proxy-based model*:

$$Y = \beta_1^{proxy} G + \beta_2 X + \epsilon. \qquad \text{(proxy-based model)}$$

As observed in prior studies, researchers use different kinds of variables to define X. By emphasizing that the chosen controls must function as proxy for Z, we are able to differentiate between two implicit strategies.

First, the researcher might choose a control variable because it is assumed to be affected by Z; call this type-A proxy. In studies of special education racial disproportionality, a common type-A proxy is prior academic performance, usually measured as a composite of students' test scores prior to disability-identification decisions (e.g., Fish 2019; Hibel et al. 2010; Kincaid and Sullivan 2017; Morgan et al. 2012, 2015, 2020; Shifrer 2018; Sullivan and Bal 2013). Intuitively, the idea is that any existing racial differences in disability would manifest in racial differences in one dimension of learning, such as a test score. Then, a test score would serve as a proxy for disability because disability is assumed to be a determinant of lower test scores.

Second, the researcher might choose a control variable because it is assumed to directly affect Z; call this type-B proxy. In studies of special education racial disproportionality, a common type-B proxy is a measure of socioeconomic status (e.g., Fish 2019; Hibel et al. 2010; Kincaid and Sullivan 2017; Morgan et al. 2012, 2015, 2020; Shifrer 2018; Sullivan and Bal 2013). Intuitively, the idea is that socioeconomic disadvantages capture factors, such as exposure to environment toxins during early childhood, that can cause disabilities. Then, a socioeconomic indicator could serve as a proxy for disability because socioeconomic disadvantages are assumed to be determinants of disability.

At first, choosing a proxy based on theoretically informed assumptions about whether it causes (or is caused by) the latent variable of interest might seem like a reasonable idea. Indeed, the traditional understanding is that the primary concern around proxy X is that it should be free of omitted-variables; that is, assuming that X follows best variable-selection practices,[6] as long as there is no omitted-variable bias, then the estimate of selection-contributing inequality will be accurate. In the context of our example, this traditional understanding suggests that the inclusion of a control variable, such as test scores, as an additional proxy for disability could be problematic only if such variable has measurement error (e.g., if the scoring of exams is vulnerable to racial bias) or if such variable is a post-treatment variable (e.g., test scores measured after disability identification decisions) (Fish et al. 2024). If not for these issues, the inclusion of an additional variable is beneficial as it might help with the central concern of omitted-variable bias. However, as we will demonstrate below with the correlated proxy problem, estimates of selection-contributing inequality can be biased even if these well-known conditions of best practices for variable selection and no omitted-variable bias are met.

Let us explicitly outline this traditional understanding through DAGs. Assume that X

[6]Though we see no mention of this in the special education literature, following best practices, X should also not contain group-level measurement error or a post-treatment effect.

does not have measurement error, does not captures a post-treatment effect, and that it captures all group-level differences with respect to Z (a no-omitted-variable bias assumption). Then, the data-generating process for type-A and type-B proxies can be defined by Figures 3A and 3A.

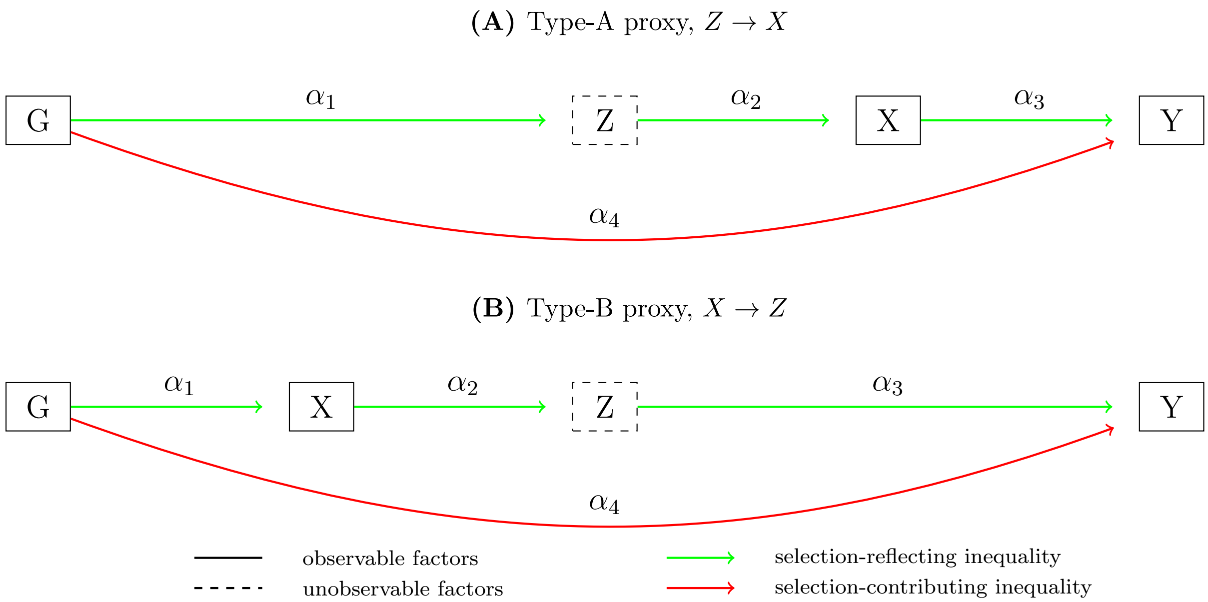


Figure 3: Two possibilities of chosen proxies and the implicit data-generating processes to appropriately assess selection-contributing inequality

In each of these cases, *the effect of G on Y that passes through Z is the same as the effect of G on Y that passes through the chosen control variable, X.* Therefore, when controlling for X in the proxy-based model the researcher is fully controlling for criteria-related group-level differences and, hence, the proxy-based model and ideal model will yield identical results.

Omitted-variable bias concerns

Briefly, we can illustrate the omitted variable bias concern; that is, the bias that occurs when the variable (or set of variables) X fails to fully capture group-level criteria-related differences. To illustrate, let V represent the unobserved variable (or set of variables) omitted from the proxy-based model. From a DAG representation, this means that, in contrast to

Figures 3A and 3B, the true data-generating processes might actually follow Figures 4A and 4B, below.

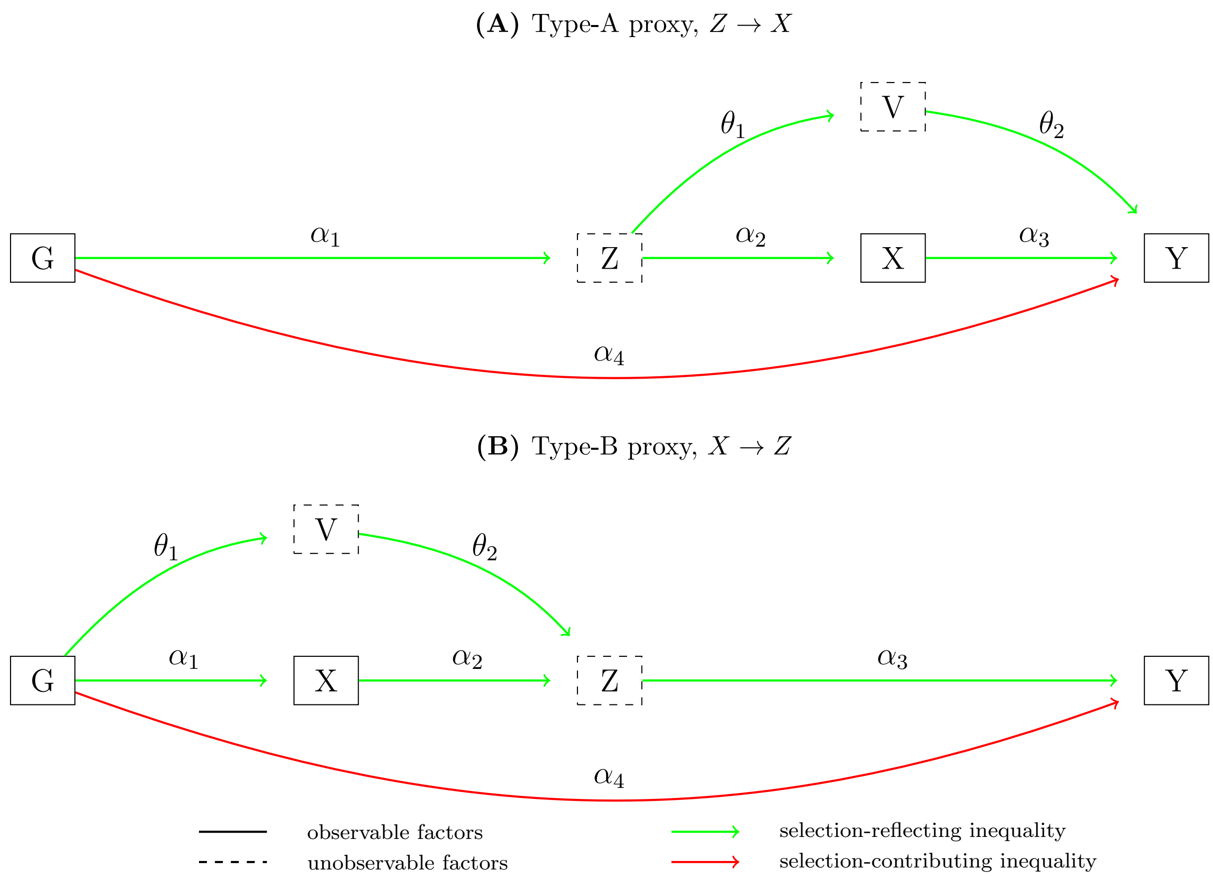


Figure 4: Omitted variable bias for type-A and type-B proxies.

To make sense of this figure, let us revisit our running example. Definitionally, a specific learning disability is unexplained low test score performance (IDEA-Sec.300.309 2006); therefore, the test score alone might not be a sufficient proxy for disability. If, conditional on test scores, there exist group-level differences with respect to exclusionary factors — such as educational opportunity and other disabilities— these differences must be accounted for. Consider cases where researchers use test scores as type-A proxies, as illustrated in Figure 4A, where X is exemplified by a test score variable. Let V represent an exclusionary factor, such as a visual impairment, and allow V to (conditional on X) differ by race, meaning that Black students are more at risk than White students to have visual impairments. By omitting V, unobserved differences between Black and White students in Z (e.g., visual impairments) persist as unmeasured confounders through θ_2 and result in omitted variable bias. The path θ_1 represents residual unobserved differences in Z as represented by V, whereas θ_2 represents racial differences in gatekeepers' selection responses to those differences. If paths θ_1 and θ_2 exist, then omitting V from the proxy-based model will omit criteria-related differences and, therefore, will yield an inaccurate estimate of selection-contributing inequality.

Now consider the type-B proxy, shown in Figure 4B, where X is exemplified by a family economic disadvantage. Certainly, family economic disadvantage is not deterministic of disability Z. It is likely that there are other structural inequalities, V, such as racial differences in exposure to lead (e.g., Lanphear et al. 1996), that (net of economic disadvantage) remain correlated with race and disability. As before, path θ_1 captures residual differences across G in the latent variable Z that affect Y, and θ_2 captures residual differences in the influence of Z on Y. If paths θ_1 and θ_2 are non-zero, then, again, omitting V from the proxy-based model will omit criteria-related differences leading to an inaccurate estimate of selection-contributing inequality.

This issue of omitted-variable bias governs much of the empirical quantitative literature on selection-contributing inequality. In fact, attempting to mitigate such concerns, special education studies (e.g., Fish 2019; Hibel et al. 2010; Kincaid and Sullivan 2017; Morgan et al. 2012, 2015, 2020; Shifrer 2018; Sullivan and Bal 2013) often introduce various other controls, such as indicators of health (e.g., birth weight); indicators of early childhood experiences (e.g., frequency of reading at home, frequency of visits to health centers, center care, birth order and parental background information); and additional measures of academic background (e.g., attendance and suspension rates). Even after the inclusion of all these controls, concerns of omitted-variable bias remain, informing sensitivity-based strategies to assess the robustness of estimates to additional unobserved confounders (Morgan 2021).

While omitted-variable bias is certainly an important methodological concern, we show that accurate estimates of selection-contributing inequality can persist *even when omitted-variable bias is not present* due to the correlated proxy problem, detailed in the next section.

5 The correlated proxy problem

To start, maintaining our prior assumption that X does not contain measurement error or a post-treatment effect,let us consider a hypothetical scenario in which the condition of no omitted-variable bias is satisfied, as shown in Figures 3A and 3A. Finally, as noted above, for clarity of exposition, our examples consider contexts in which X is defined by a single variable.

The issue is that the failure to capture group-level differences with respect to Z (omitted-variable bias) is not the only way controls can create false impressions about selection-reflecting inequality. The specific form of the correlated proxy problem depends on the kind of proxy chosen by the researcher (type-A or type-B), and the DAGs in Figures 5A and 5B illustrate the correlated proxy problem for both types. Importantly, note that the DAGs presented in this Figure modify the previously assumed data-generating processes (Figures 3A and 3B) while keeping the assumption of no omitted-variables bias.

First, let us focus on type-A proxy, Figure 5A. In this revised data-generating process, we introduce path γ, which defines an effect of G and X that does not passes through Z. In less abstract language, this means that there might exist some ex ante factors that lead X to vary across group membership but that do not explain why Z varies across groups. The effect of G on Y that passes through Z is no longer the same as the effect of G on Y that passes through X. Consequently, the effect of X on Y (α_3) can now capture both a selection-contributing and selection-reflection inequality, what we highlight by changing its color to blue. Hence, when we estimate the proxy-based model, we are not controlling only for criteria-related differences (paths α_1, α_2 and α_3) but also for criteria-unrelated ones (path γ).

To illustrate, let us go back to our running example of racial disproportionality in special

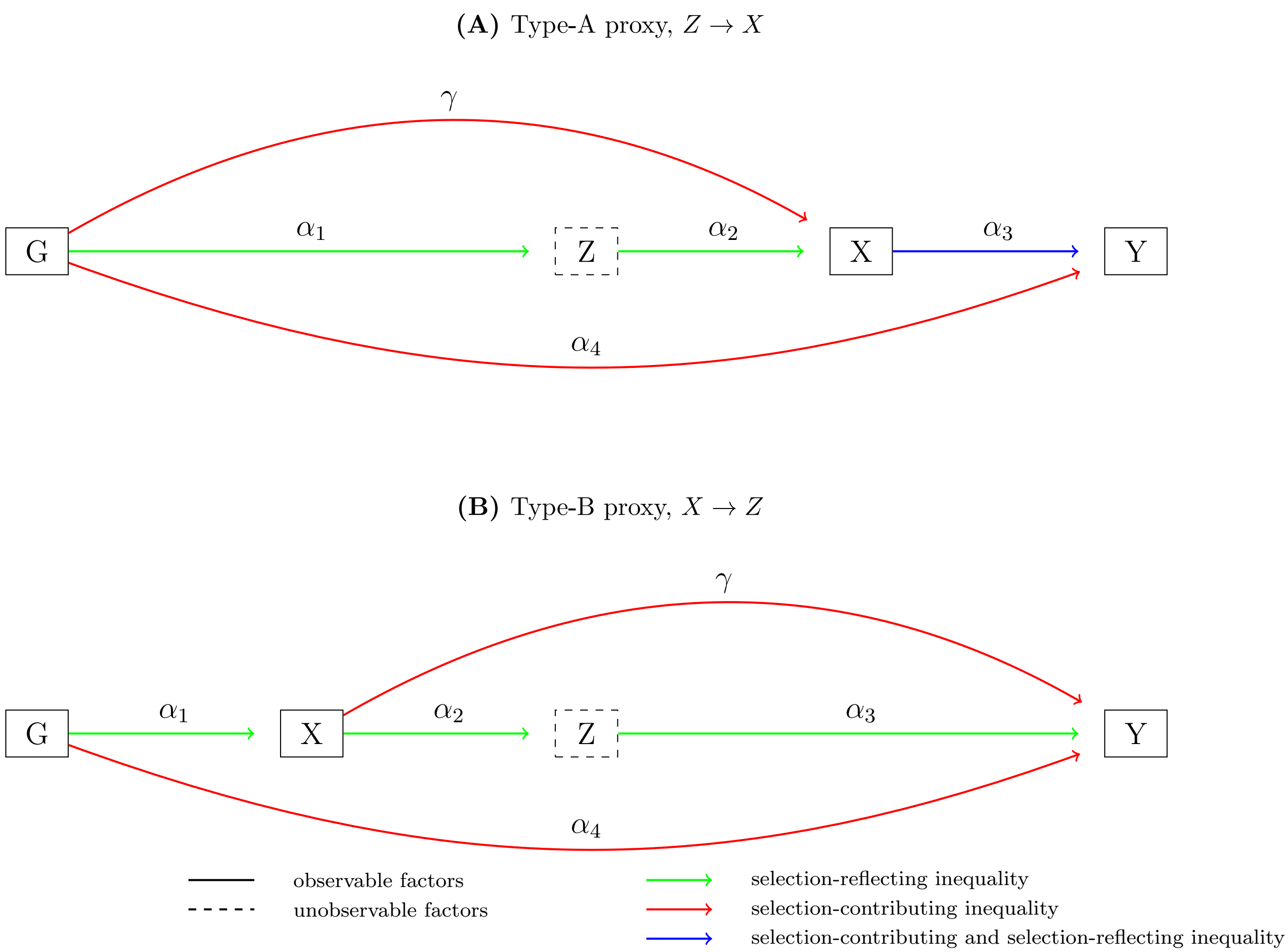


Figure 5: The correlated proxy problem under two types of chosen proxies

education for students with SLD. In this case, Z, represents whether a student has developed a learning disability prior to special education identification decisions. Recall that IDEA defines a learning disability as an academic difficulty that cannot be attributed to other disabilities or the lack of educational opportunities. The correlated proxy problem can arise in this case if the chosen type-A proxy (exemplified by test scores) captures Black-White differences in educational opportunities. If that is the case, a model that controls for test scores is implicitly controlling for factors that are, as represented by Figure 5A, distinct from a learning disability, thereby biasing the estimate of interest.

Consider, for instance, the fact that parental practices, such as reading habits and the nature of parent-child interactions, play a central role in the development of children's cognitive skills (Britto et al. 2017; Hemmerechts et al. 2017; Imhof et al. 2023; Mendive et al. 2017). Due to historical and structural racial inequalities, different-race families might have different financial and cultural resources that translate into differential parental practices and, ultimately, into racial differences in test scores (path γ). The problem, therefore, is that test scores might capture these unequal learning opportunities that, by definition, are not legitimate causes of learning disability.

Controlling for test scores in this regression would lead us to compare students at the same achievement level while failing to distinguish between those who are low-achieving due to a learning disability (the necessary covariate to identify selection-contributing inequality) versus those who are low-achieving due to lack of educational opportunity (the component of selection-reflecting inequality). This statistical adjustment would therefore underestimate true selection-contributing racial disparities in special education access by incorrectly attributing all achievement differences to selection-reflecting factors. In other words, by controlling for test scores that capture both legitimate disability-related differences and illegitimate opportunity-related differences, we risk masking the selection-contributing component of racial disparities that stems from differential educational opportunities.

Second, consider type-B proxy, Figure 5B. In the revised data-generating process, we introduce path γ, which posits that the extent to which group-level differences in X matter for Y do not fully operate through Z. The effect of G on Y that passes through Z is no longer the same as the effect of G on Y that passes through X. While the effect of G on Y that passes through Z only captures selection-reflecting inequality (paths $\alpha_1, \alpha_2, \alpha_3$), the effect of G on Y that passes through X captures both selection-reflecting (paths $\alpha_1, \alpha_2, \alpha_3$) and selection-contributing inequalities (path γ). Therefore, when we estimate the proxy-based model, we are not solely controlling for criteria-related differences (paths α_1, α_2 and α_3) but also for criteria-unrelated ones (path γ).

To illustrate, let socioeconomic background be the type-B proxy the researcher uses to identify selection-contributing inequality for group-differences in identification of students with SLDs. It is possible that a student's socioeconomic background matters for special education receipt in ways that are not explained by learning disability. For instance, consider the case of parental involvement. Through involvement in the school, parents can exert pressure on teachers and administrate — as well as can teach their children to exert such pressure on their own (Calarco 2018) — to customize educational experiences, influencing, for instance, whether their children are assigned to various educational services, such as gifted and talented programs, advanced-coursework and special education (Cowhy et al. 2024; Lareau 2011; Lewis and Diamond 2015; Lewis-McCoy 2014; Tyson 2011). One's ability to successfully influence school practices is highly correlated with socioeconomic background (path γ), since social position conveys cultural and social capital resources necessary for effective parental influence (Calarco 2018; Lareau 2011; Lewis-McCoy 2014). Therefore, a model that controls for a socieconomic indicator might account for racial differences in special education receipt that are unrelated to learning disability, hence, biasing the estimate of selection-contributing inequality.

Controlling for socioeconomic background in this regression would lead us to compare students of similar socioeconomic status while failing to distinguish between differences in special education receipt that stem from learning disability status (the necessary covariate to identify selection-contributing inequality) versus those that stem from parents' differential ability to advocate for services (the component of selection-reflecting inequality). This statistical adjustment would therefore underestimate true selection-contributing racial disparities in special education access by incorrectly attributing all socioeconomic differences to selection-reflecting factors. In other words, by controlling for socioeconomic status that captures both legitimate disability-related differences and illegitimate advocacy-related differences, we risk masking the selection-contributing component of racial disparities that stems from differential ability to navigate the special education system.

One might point out that, in practice, to minimize concerns of omitted-variable bias, studies report a series of proxy-based models, estimating models that include several proxy variables. It is reasonable to ask: how we can make sense of the correlated proxy problem in a context where X is defined by multiple variables? Figures B.1A and B.1A in Appendix B extend our conceptual framework to show how a correlated proxy problem can arise when the proxy-based model is defined by two control variables. Following the DAGs in Figures 4A and 4B, we assume that two variables (X and V) are needed so that the data-generating process meets the assumption of no-omitted-variable bias. Therefore, the inclusion of V into the model is necessary to satisfy the no omitted-variable bias assumption. However, it is unlikely that such an additional variable can help mitigate the extent to which X brings about concerns regarding a correlated proxy problem (paths γ in Figures 5A and 5B). In fact, such additional variable can introduce further backdoor paths (paths γ_2 in Figures B.1A and B.1B), creating additional possibilities of bias via a correlated proxy problem. This exercise shows that balancing concerns between omitted-variable bias and potential further correlated proxy problems is a fundamental statistical challenge of regression-based frameworks assessing selection-contributing inequality.

A formalization of the correlated proxy problem

Given the intuition for the correlated proxy problem developed so far, we now formally derive how it might introduce bias in the estimand of interest. The form of the bias depends on the type of proxy chosen by the researcher. Therefore, we separately discuss the biases introduced by type-A and -B proxies.

Type-A proxy

Informed by Ding and Miratrix (2015), we can represent the data-generating process from Figure 5A's DAG through the following linear simultaneous equations model (LSEM):

$$
\begin{cases}
\text{G} & \sim B(p) \\
\epsilon_1, \epsilon_2, \epsilon_3 & \overset{\text{ind}}{\sim} N(0,1) \\
Z & = \alpha_1 G + \epsilon_1 \\
X & = \gamma G + \alpha_2 Z + \epsilon_2 \\
Y & = \alpha_3 X + \alpha_4 G + \epsilon_3
\end{cases}
\tag{1}
$$

In this system, G represents a binary variable distributed with probability p according to a Bernoulli distribution. For mathematical simplicity and generality, let Z, X and Y be continuous variables. Z is the (standardized) probability that individual i meets the selection criteria; X is the (standardized) proxy variable (or composite of proxy variables) for Z; and Y is the (linear) probability that individual i is assigned to the program. Parameters define linear relationships between variables. The γ parameter captures the problematic path that brings about the correlated proxy problem as shown in Figure 5B's DAG. If $\gamma = 0$, the structure of the data-generating process will match exactly what we would need to assume for the proxy-based model to appropriately assesses selection-contributing inequality (Figure 3A's DAG).

Based on this data-generating process, we can now derive the values of $\hat{\beta}_1^{raw}$, $\hat{\beta}_1^{ideal}$ and $\hat{\beta}_1^{proxy}$. See Appendix A for derivations.

$$\hat{\beta}_1^{raw} = \alpha_1\alpha_2\alpha_3 + \alpha_3\gamma + \alpha_4; \tag{2}$$

$$\hat{\beta}_1^{ideal} = \alpha_3\gamma + \alpha_4; \tag{3}$$

$$\hat{\beta}_1^{proxy} = \alpha_4. \tag{4}$$

These results help us formally illustrate the arguments outlined above. First, we highlight that the estimate from the raw model, Eq. 2, depends on all of the paths through which group G matters for selection into program Y, paths α_1, α_2, α_3, α_4 and γ in Figure 5A's DAG. This illustrates mathematically that the raw model cannot appropriately assess selection-contributing inequality because it captures the multiple processes through which program disproportionality emerges.

Second, and most importantly, these results show that, despite assuming a data-generating that allows the proxy-based model to meet the condition of no omitted-variable bias, we observe a difference between the $\hat{\beta}_1^{ideal}$, Eq. 3, and $\hat{\beta}_1^{proxy}$, Eq. 4. This result shows that the proxy-based model does not necessarily provide an accurate estimate of selection-contributing inequality (what is done by the ideal model). In fact, the difference between $\hat{\beta}_1^{ideal}$ and $\hat{\beta}_1^{proxy}$ captures the level of bias in the proxy-based estimate introduced by the correlated proxy problem (5' DAG). Formally, the correlated proxy problem for a type-A proxy can be written as:

$$CPP = \hat{\beta}_1^{ideal} - \hat{\beta}_1^{proxy} = \gamma\alpha_3. \tag{5}$$

From this exercise, we learn that further conditions — in addition to no omitted-variable bias and that the proxy does not contain measurement error or a post-treatment effect — need to hold for a proxy-based model (that is based on a type-A proxy) to accurately estimate selection-contributing inequality, $CPP = 0$. Specifically, it must be the case that at least one of the paths γ or α_3 in Figure 5A's DAG is equal to 0.

If $\alpha_3 = 0$, then we have a data-generating processes in which the proxy does not explain group-level differences in program enrollment. Note that if that is the case, then all models will produce the same estimate, $\hat{\beta}_1^{raw} = \hat{\beta}_1^{ideal} = \hat{\beta}_1^{proxy} = \alpha_4$, suggesting that α_4 is the only path through which group membership shapes Y. This result, while mathematically true, should be of little relevance to practice since, in reality, the chosen proxy should, by definition, explain group-level differences in program enrollment. In fact, α_3 is reported by existing empirical estimates (i.e., it is the coefficient on X in the proxy-based model) and is rarely (if ever) reported to be 0.

Give that α_3 is empirically known and, by assumption, $\alpha_3 \neq 0$, then the extent to which the proxy-based model is vulnerable to the correlated proxy problem depends on γ. Intuitively, the γ path captures the extent to which the chosen proxy is correlated with group-level differences that are unrelated to Z. If such correlation is 0, then $\hat{\beta}_1^{ideal} = \hat{\beta}_1^{proxy}$ and, hence, there will be no correlated proxy problem.

This result differs from standard omitted variable bias in several important ways. First, in typical omitted variable bias, we worry that excluding a variable leads to biased estimates because that variable affects both the treatment and outcome. Here, even when we meet the standard conditions for avoiding omitted variable bias, we still face the correlated proxy problem because the proxy itself (X) captures group differences through two distinct pathways - one through the latent characteristic (Z) and another directly (γ).

Second, while omitted variable bias typically suggests adding more controls would improve our estimates, the correlated proxy problem shows that controlling for certain variables can actually introduce bias even when they appear to be relevant controls. This occurs because the proxy captures both legitimate differences that should be controlled for (via Z) and illegitimate differences that should not be controlled for (via γ).

Third, unlike omitted variable bias which can often be addressed through better measurement or additional controls, the correlated proxy problem presents a fundamental identification challenge. As shown in equation 5, the bias depends on γ, which captures the extent to which group differences in the proxy are unrelated to the latent characteristic we actually want to measure. Since γ is inherently unobservable and α_3 is typically non-zero in practice, we cannot eliminate this bias simply by adding more variables to the model.

Overall, this derivation shows that assessing whether a proxy-based model based on a type-A proxy which meets the well-know condition of no omitted-variable bias (and follows best practices of not capturing measurement error or post-treatment effects) can accurately assess selection-contributing inequality depends on the unknown nature of γ.

Type-B proxy

Now, let us consider the case where the researcher chooses a proxy under the assumption that the proxy is a cause (instead of an effect) of the latent concept of interest, Z — what we have called a type-B proxy. We can represent the data-generating process from Figure 5B's

DAG through the following linear simultaneous equations model (LSEM):

$$\begin{cases} \text{G} & \sim B(p) \\ \epsilon_1, \epsilon_2, \epsilon_3 & \overset{\text{ind}}{\sim} N(0,1) \\ X & = \alpha_1 G + \epsilon_1 \\ Z & = \alpha_2 X + \epsilon_2 \\ Y & = \alpha_3 Z + \alpha_4 G + \gamma X + \epsilon_3 \end{cases} \tag{6}$$

The definition of variables G, Z, X and Y follow the LSEM 1. Now, however, the γ parameter captures the problematic path that brings about the correlated proxy problem as shown in Figure 5B's DAG. If $\gamma = 0$, the structure of the data-generating process will match exactly what we would need to assume for the proxy-based model to appropriately assesses selection-contributing inequality (Figure 3B's DAG).

The values of $\hat{\beta}_1^{raw}$, $\hat{\beta}_1^{ideal}$, $\hat{\beta}_1^{proxy}$ and of the correlated proxy problem (CPP) can now be written as follows. See Appendix A for derivations.

$$\hat{\beta}_1^{raw} = \alpha_1\alpha_2\alpha_3 + \alpha_1\gamma + a_4; \tag{7}$$

$$\hat{\beta}_1^{ideal} = \frac{\alpha_1\gamma + \alpha_2^2\alpha_4 + \alpha_4}{\alpha_2^2 + 1}; \tag{8}$$

$$\hat{\beta}_1^{proxy} = \alpha_4; \tag{9}$$

$$CPP = \hat{\beta_1}^{ideal} - \hat{\beta_1}^{proxy} = \frac{\alpha_1\gamma}{\alpha_2^2 + 1}. \tag{10}$$

Similar to what have seen in the analysis of a type-A proxy, we observe that the estimate from the raw model, Eq. 7, depends on all of the paths through which group G matters for selection into program Y, paths α_1, α_2, α_3, α_4 and γ and, thus, that the raw model captures all the processes (selection-contributing and -reflecting) through which program group-level disproportionality emerges.

Further, we see, again, that $\hat{\beta_1}^{ideal} \neq \hat{\beta_1}^{proxy}$ and, thus, that the proxy-based model does not necessarily provide an accurate estimate of selection-contributing inequality. The form of the correlated proxy problem, in this case, is captured by Eq. 10, which depends on α_1, α_2 and γ. Now, we will have no correlated proxy problem if at least one of paths γ or α_1 are 0.

A data-generating process where $\alpha_1 = 0$ implies that there are no group-level differences in the chosen proxy. The idea of a proxy is, by assumption, to capture group-level differences with respect to the latent concept of interest, Z. Therefore, as empirically observed, researchers choose a proxy such that $\alpha_1 \neq 0$. Note that results here indicate that, assuming that $\gamma \neq 0$, the more X differs across groups (as captured by α_1) the stronger the correlated proxy problem will be.

Given that α_1 is empirically observed, the level of the correlated proxy problem depends on two unknowns, α_2 and γ. Parameter γ specifies that the more X explains group-level differences in program enrollment through paths that are unrelated to criteria-related differences the stronger the correlated proxy problem will be. Further, parameter α_2 captures that the higher the relationship between the chosen proxy and the latent concept of interest, Z, the weaker the correlated proxy problem will be.

Like the Type-A case, the Type-B correlated proxy problem presents a distinct challenge from omitted variable bias. Even when we satisfy traditional conditions for avoiding omitted variable bias, the proxy variable (X) affects program enrollment (Y) through two distinct mechanisms: one mediated by the latent characteristic (Z) through paths α_1 and α_2, and another direct effect through path γ.

The key distinction is that in standard omitted variable bias, adding controls helps reduce bias by accounting for confounding relationships. However, in the Type-B correlated proxy problem, controlling for the proxy variable can actually introduce bias even when it appears to be a relevant control. This occurs because the proxy's relationship with program enrollment combines both legitimate effects (operating through the latent characteristic Z) and illegitimate effects (operating directly through γ).

Moreover, as shown in equation 10, the magnitude of this bias depends on both γ and α_2—parameters that cannot be directly observed. Since α_1 is typically non-zero by design (as we choose proxies specifically because they vary across groups), we cannot eliminate this bias through conventional approaches to addressing omitted variable bias. This creates a fundamental identification challenge that persists even when standard econometric conditions are met.

Overall, this derivation shows that assessing whether a proxy-based model based on a type-B proxy which meets the well-know condition of no omitted-variable bias (and follows best practices of not capturing measurement error or post-treatment effects) can accurately assess selection-contributing inequality depends on the unknown natures of α_2 and γ.

6 Empirical illustration of the correlated proxy-problem

So far, we have argued that the correlated proxy problem can lead proxy-based models to (in theory) produce biased estimates of selection-contributing inequality even when models are not vulnerable to common concerns of omitted-variable bias. Now, we apply the framework above to an empirical context and illustrate how this methodological issue is not simply a theoretical possibility and that it can, in fact, be meaningful to existing academic and policy debates.

Here we empirically examine our running example of racial disproportionality in special education. We consider data from the National Center for Education Statistics (NCES) 1998 Early Childhood Longitudinal Study, Kindergarten cohort (ECLS-K, public-use version) to assess how the receipt of special educational services (defined as the receipt of an IEP) for any disability category (program Y) varies across Black and White students (group membership G, where "Whites" is the reference category). We chose the ECLS-K for this empirical exercise because it is the most frequently used data source in quantitative studies of racial disproportionality in education. We note, however, that the ECLS-K has important limitations for the analyses of special education — e.g., it has small sample sizes, making it difficult to achieve statistically significant inferences; its weight-adjusted estimates of special education receipt rates do not closely follow what national-level data suggests (Skiba et al. 2016a). Despite these limitations, the ECLS-K provides the best accessible student-level data for a national-level assessment of special education disproportionality. Therefore, by relying on the ECLS-K, we can demonstrate how relevant the correlated proxy problem is to the existing literature.

According the framework proposed here, to derive appropriate conclusions around selection-contributing inequality, researchers should, first, clearly articulate what is the normative selection criteria governing selection, variable Z. Special education services for students with an SLD are reserved for students with a specific learning disability—by definition, not caused by other disabilities or exclusionary factors.[7]

Next, the researcher should compare group-level differences in program enrollment rate net of differences in how individuals meet the selection criteria. This means that, ideally, we should estimate Black-White differences in IEP receipt rates net of Black-White differences in incidences of latent disability (prior to selection decisions).

However, the latent concept of disability is unobservable to the researcher. Existing quantitative analyses of the ECLS-K adopt what we have been calling a proxy-based strategy, where scholars estimate Black-White differences in IEP receipt rates net of various controls. To illustrate the correlated proxy problem for a type-A and type-B proxy, we focus on two most common chosen variables throughout our examples above. First, a composite of students' test scores measured at kindergarten entry (and, thus, before special education decisions take place). In this case, as outlined above, the implicit assumption is that test scores could appropriately operationalize disability because disabilities are assumed to manifest into lower test scores. Thus, test scores exemplify a type-A proxy for disability. Second, a composite measure of students' socioeconomic background measured at kindergarten entry. The implicit assumption in this case is that economic disadvantages can create conditions (such as exposure to lead, poor air quality) that can lead to disabilities. Then, socioeconomic status exemplifies a type-B proxy for disability.

[7] In the public-use ECLS-K we can only observe a binary indicator for whether a student has an IEP or not, and therefore must treat this omnibus disability category as if it also adheres to (IDEA 2004) specifications about the definition of disability and exclusionary factors. Given that SLD is the most common disability category, this treatment of the IEP variable seems reasonable for illustration purposes.

Data

We use data from ECLS-K's 6th round of data collection (grade 5). To closely dialogue with existing research, our sample-selection and data-handling practices follows prior quantitative assessments of racial disproportionality in special education (e.g., Hibel et al. 2010). Of the students that had available data in the 6th round of data collection, we only consider students who had non-missing indicators of IEP receipt status on students' records abstract. Because of our focus on Black-White inequalities, we concentrate only on the subsample of Black (n = 1,007) and White (n = 5,479) students.

Following existing research, our test score measure is a standardized composite of students' math and reading scores in item-response theory exams in the first semester of kindergarten (and thus, measured before the special education decisions of interest — 6th round of data collection — take place). Further, our SES measure is based on an ECLS-K-provided composite which captures factors such as level of parental education, occupational prestige, and household income. As prior studies, we handle missing cases in the test score and SES composites by conditional multiple imputation and compute estimates using ECLS-K's student-level longitudinal weights and heteroskedasticity-consistent Huber–White standard errors (Hibel et al. 2010; Morgan et al. 2015).

Raw and proxy-based models

Given this quantitative operationalization of variables Y, G and X, we now estimate the raw model and proxy-based model defined above. Table 2 reports the results. Coefficients capture the average marginal difference of being from the "Black" racial category (as opposed to the "White" racial category). This unconditional point estimate of 0.03 matches data from the Office of Special Education Programs (U.S. Department of Education 2019) but is not statistically significant due to the ECLS-K having a small sample size, a result reported previously (e.g., Hibel et al. 2008; Kincaid and Sullivan 2017). Thus, for the sake of this empirical illustration, we do not concentrate on statistical significance and focus on the

Table 2: Raw and proxy-based models. Coefficients capture the average marginal difference of variables of interest on one's likelihood of special education education receipt.

	Raw model	Proxy-based model, Type-A proxy (X = test score)	Proxy-based model, Type-B proxy (X = SES)
(Intercept)	0.13^{***}	0.48^{***}	0.14^{***}
	(0.01)	(0.04)	(0.01)
race = Black	0.03	-0.09^{***}	0.01
	(0.03)	(0.03)	(0.03)
test score		-0.28^{***}	
		(0.02)	
SES			-0.04^{**}
			(0.01)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

magnitude of the effects. To understand how to interpret these results, consider the Raw model. A 0.03 average marginal difference of Black versus White on the probability of special education receipt means that, on average, Black students' probability of special education receipt is 3 percentage points higher than Whites'. Note that only about 10% of students in our sample receives special education services and, thus, that a 3-point percentage difference is, in fact, quite substantial. To better visualize these results, Figure 6 plots the predicted probabilities of special education receipt according to the coefficients in Table 2. For the two proxy-based models, control variables are held at their means.

The results using test score and socioeconomic proxies as covariates replicate prior studies (e.g., Hibel et al. 2008, 2010). These proxy-based models show that racial differences in prior test scores or socioeconomic status (chosen proxies for disability) can explain (at least partially) such disparities. When test score is the chosen control, we observe that Black students are, in fact, less likely than White students to receive an IEP. When socioeconomic status is the chosen control, we observe that Black students are almost just as likely as Whites

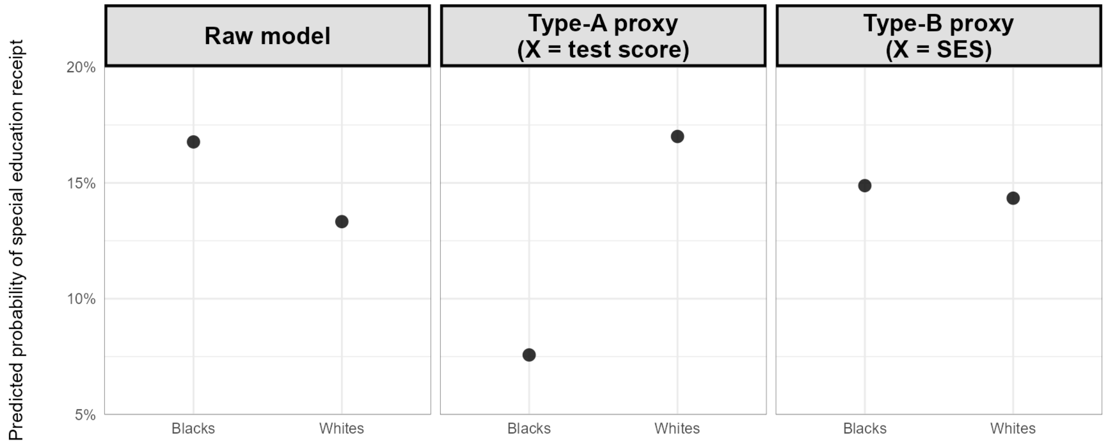


Figure 6: Predicted probabilities of special education receipt according the coefficients in Table 2. For the two proxy-based models, control variables are held at their means..

to receive an IEP. These results have been used as evidence that Black students are under-represented in special education relative to White students, once appropriate comparison groups are established via covariate adjustment (e.g., Gordon 2017; Morgan and Farkas 2015). In terms of the language we propose here, these models have been interpreted as evidence of selection-contributing inequality, meaning that Black students have been denied access to services for which they are eligible at greater rates than White students.

Now, we assess how the proxy-based models presented above can be vulnerable to the correlated proxy problem. Naturally, these models can also be vulnerable to the issue of omitted-variable bias. The purpose of this exercise, however, is to show that, in contrast to traditional understanding, omitted-variable bias is *not the only methodological issue we should be concerned about and that the proxy-based model can inaccurately assess selection-contributing inequality even if it is not vulnerable to omitted-variable bias.*

To do so, let us consider the (perhaps hypothetical) case in which the chosen proxy-based models meet the no-omitted-variable bias condition. Then, we assess how problematic paths (as defined by DAG 5A and 5A) can bring about the correlated proxy problem.

Type-A proxy (X = test score)

Our goal in this exercise is to generate a plausible range of values for γ to test the effect of introducing correlated proxies on Black-White differences in the receipt of special education services. Based on Eq. 5, the correlated proxy problem for a type-A proxy only depends on α_3 and γ. Parameter α_3 is known to the researcher as it is directly estimable through the proxy-based model. In fact, from Table 2, we know that $\alpha_3 = -0.28$. Therefore, to assess the extent to which the proxy-based estimate, $\hat{\beta^{proxy}}_1$ is robust to the correlated proxy problem, there is only one unknown, γ, which in this empirical application, captures the extent to which test scores, X, differ across students of different races, G, but similar levels of disability, Z. Thus, we can rely on empirical and theoretical insights to assume a plausible range for γ. Based on this range, we can construct an interval around the proxy-based estimate, $\hat{\beta^{proxy}}_1$, displaying the extent to which the correlated proxy problem would allow it to vary.

To define a plausible range for γ, consider that national descriptive statistics show that, on average, Black students have lower test scores than White students, at between 0.50 to 0.75 standard deviations depending on the data, year, and grade (Quinn 2015; Reardon et al. 2019a,b). Further, it is certainly the case that factors beside disability, such as racial segregation (Reardon et al. 2019b) and economic inequality (Christopher 1996), cause differences in test scores, and, thus, that racial differences in disability are unlikely to fully explain the Black-White test-score gap. Thus, we assume that for the highest plausive level of Black-White differences in disability, variable Z could fully explain the Black-White test score gap, but not reverse it. That is, it is unreasonable to believe that, conditional on disability, Black students would have higher test scores than Whites, given the multiple structural inequalities undergirding test scores. This means that we can safely assume that $\gamma \leq 0$ — i.e, conditional on disability, the Black-White test score gap is, at most, 0. To define a lower bound for γ, we assume that for the lowest plausive level of Black-White differences in disability, Z would play no role in explaining Black-White differences in test scores. Then, γ would be the same as the coefficient on G for a regression of X on G, the unconditional Black-White difference in test scores. For our sample, such a coefficient is -0.47. Therefore, we assume $\gamma \in \{-0.47, 0\}$.

To help with the interpretation of our results, note that when defined within this range, chosen values of γ imply an assumption about the percentage of the Black-White test score gap which is explained by disability. A value of $\gamma = -0.47$ implies, for instance, that the Black-White test score gap is the same regardless of whether we control for Z and, therefore, indicates that disability explains 0% of the Black-White test score gap. Conversely, a value of $\gamma = 0$ indicates that disability explains 100% of the Black-White test score gap. More formally, let d capture the Black-White test score gap conditional on disability over the unconditional Black-White test score gap. Then, $\gamma = -0.47$ implies $d = 0\%$ and $\gamma = 0$ implies $d = 100\%$.

Based on this chosen range, and based on the fact that the ideal estimate (Eq. ideal model) is simply the sum of ($\hat{\beta}_1^{proxy}$) and the correlated proxy problem ($CPP = \alpha_3\gamma$), we can compute a range of plausible values for the ideal estimate ($\hat{\beta}_1^{ideal}$), thereby assessing how the correlated proxy problem might bias the proxy-based estimate. Figure 7 presents the results of this exercise. Given our assumptions about the nature γ, the correlated proxy-problem will be non-negative, and the ideal estimate will fall somewhere between the raw gap and the proxy-based estimate.

To help interpret our analyses, Figure 7 shows how the ideal estimate varies with assumed values of γ. We highlight what these values of γ imply about the percentage, d, of the Black-White test score gap explained by disability, Z. This helps readers assess which values are more plausible.

Setting $\gamma = 0$ implies that disability explains 100% of the Black-White test score gap. For context, socioeconomic status explains about 29% of the Black-White achievement gap in our data, consistent with prior research (Reardon et al. 2019a). If disability explains 100% of the Black-White test score gap, this implies a magnitude for the disability to non-

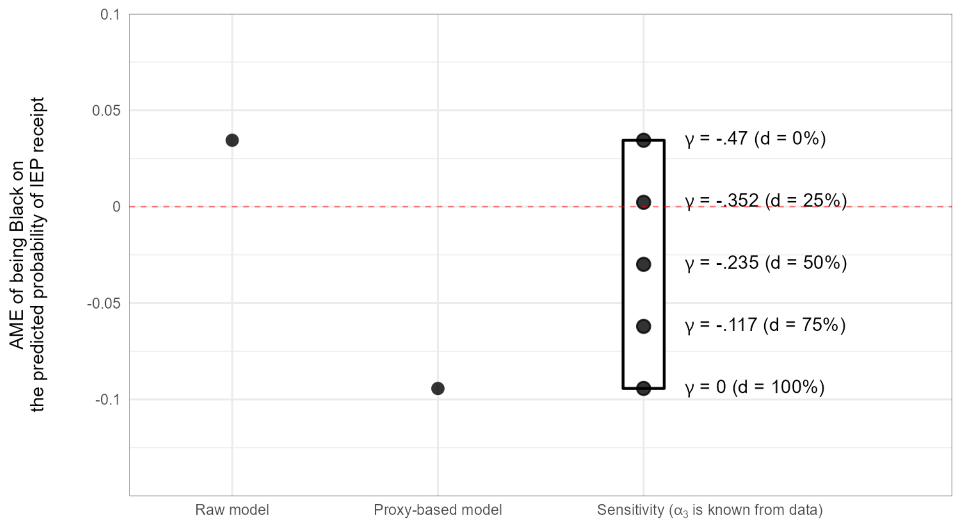


Figure 7: Robustness of the proxy-based model (for type-A proxy, X = test score) estimated in Table 2 to the correlated proxy problem. The sensitivity band defines the range of possible values for the ideal estimate for $\gamma \in \{-0.47, 0\}$. Parameter α_3 is known from data (= -0.28). To facilitate interpretation, we highlight the sensitivity estimates where $\gamma \in \{-0.47, -0.352, -0.235, -0.117, 0\}$. Values of γ imply an assumption about the percentage, d, of the Black-White test score gap that is explained by disability, Z.

disability gap, assuming other parameters. For example, assigning population parameters to the (unobserved) disability rate and Black-White disability gap (15% overall disability rate and a very conservative 22.5% to 7.5% Black-White disability gap) entails a disability-related achievement gap of more than 4 standard deviations (SD), a value nearly 6.5 times larger than current estimates of 0.62 SD (Gilmour et al. 2019). Alternatively, if we now allow the disability-related achievement gap to be 0.62 SD, then the true (unobserved) disability rate gap for Black and White students would need to be 117%, a number that is impossible. Even if disability explains just 25% of the achievement gap ($\gamma = -0.352$), the implied disability achievement gap would be 1.17 SD, or the Black-White disability rate gap would be 28.2%. See Appendix C for details on these calculations.

These calculations mean that using test scores to control for Black-White differences in IEP receipt very likely exaggerates the magnitude (and perhaps mis-estimates the sign) of selection-contributing inequality by conflating structural factors affecting the Black-White achievement gap with unobserved differences in disability probability between Black and White students. While the true disability rate, racial differences in disability rates, and achievement levels of students with disabilities are unknown, our calculations use conservative assumptions that, if adjusted, would imply even larger disability gaps.

Type-B proxy (X = SES)

Based on Eq. 10, the correlated proxy problem for a type-B proxy depends on α_1, α_2 and γ. Parameter α_1 is known to the researcher as it captures the average group-level differences with respect to X. In our data, $\alpha_1 = -0.74$.

To assess the extent to which the proxy-based estimate, $\hat{\beta^{proxy}}_1$, is robust to the correlated proxy problem, there are, therefore, two unknowns, α_2 and γ. Then, as done for the empirical assessment of type-A proxy in the section above, we can rely on empirical and theoretical insights to assume plausible ranges for α_2 and γ.

Parameter α_2 captures the correlation between the chosen proxy, socioeconomic status, and the latent concept of interest, disability. Under the theoretical assumption articulated above, that socioeconomic disadvantages lead to factors, such as exposure to environmental toxins and family stress, that, ultimately, influence the development of disabilities, it follows that $\alpha_2 \leq 0$. To provide a lower bound for α_2, we suppose that, at most, the absolute value of the correlation of socioeconomic status and disability is equal to the correlation between socioeconomic status and test scores, which, in our data is equal to 0.22. Therefore, we assume $\alpha_2 \in \{-0.22, 0\}$.

Parameter γ captures the extent to which X explains group-level differences in program enrollment through pathways that are unrelated to disability. As proposed in the discussion of Figure 5B's DAG, this path can arise because parents' ability to influence gatekeepers decisions depend on cultural and social capital resources which come together with socioeconomic advantages (Calarco 2018; Lareau 2011; Lewis-McCoy 2014). Therefore, net of disability, socioeconomic status should have a non-positive effect on special education receipt ($\gamma \leq 0$).

To provide a lower bound for γ, recall the Y equation in LSEM 6:

$$Y = \alpha_3 Z + \alpha_4 G + \gamma X + \epsilon_3. \tag{11}$$

Note that if X were to explain group-level differences in program enrollment through pathways that fully operate through Z (which implies $\alpha_3 = 0$), we could rewrite the Y equation as

$$Y = \alpha_4 G + \gamma X + \epsilon_3. \tag{12}$$

Then, γ would be, simply, the effect of X on Y conditional on G. From our data, we know that this value is -0.04. We propose -0.04 to be the lower bound for γ and, thus, $\gamma \in \{-0.04, 0\}$.

To help with interpretation, note that, when defined within this range, assumed values for γ capture how much of the effect of X in explaining group-level differences in program enrollment operate through Z. For instance, $\gamma = 0$ implies that all (100%) of the effect of X in explaining group-level differences in program enrollment operate through Z. Conversely, $\gamma = -0.04$ implies that none (0%) of the effect of X in explaining group-level differences in program enrollment operates through Z. More formally, let d the percentage of the effect of X on group-level differences in enrollment which operate through Z. Then, $\gamma = 0$ implies $d = 100\%$ and $\gamma = -0.04$ implies $d = 0\%$.

Based on this proposed range, we can construct an interval around the proxy-based estimate, $\hat{\beta^{proxy}}_1$, displaying the extent to which the correlated proxy problem would allow it to vary. Figure 8 presents the results of this exercise. For easier visualization, we plot sensitivity estimates for the chosen range of γ while keeping $\alpha_2 = -0.22$ (the strongest possible assumed correlation between disability and socioeconomic status). Figure B.2 in Appendix B shows that other values of α_2 for $\alpha_2 \in \{-0.22, 0\}$ have only a minor effect on the sensitivity estimates displayed here. As before, results illustrate that, given assumptions, plausible values for the true estimate will fall within the raw and proxy models.

To help interpret our analyses, Figure 8 shows how the ideal estimate varies for specific assumed values of γ. We also highlight what these values of γ imply about the percentage,

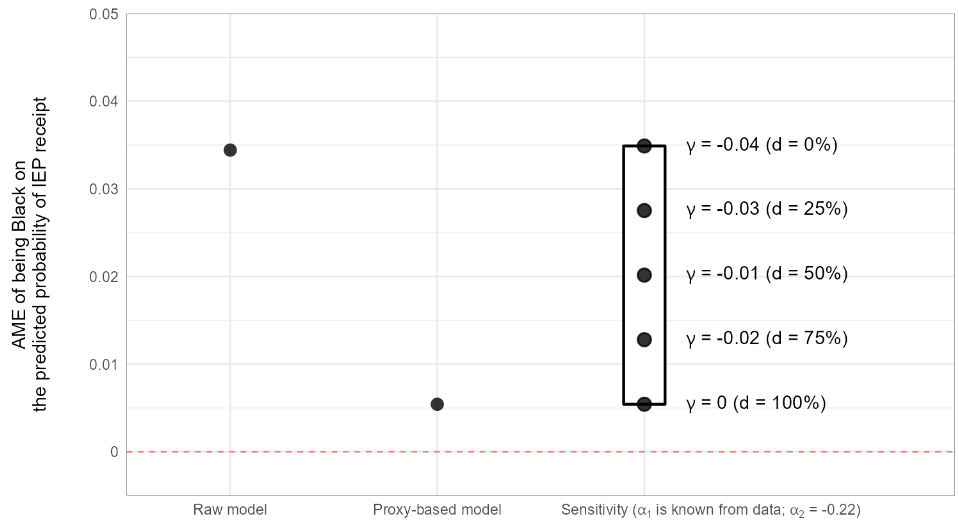


Figure 8: Robustness of the proxy-based model (for type-B proxy, X = SES) estimated in Table 2 to the correlated proxy problem. The sensitivity band defines the range of possible values for the ideal estimate for $\gamma \in \{-0.04, 0\}$. Parameter α_1 is known from data (= -0.74) and $\alpha_2 = -0.22$ (see Figure B.2 in Appendix B for sensitivity estimates where $\alpha_2 \in \{-0.22, 0\}$). To facilitate interpretation, we highlight the sensitivity estimates where $\gamma \in \{-0.04, -0.03, -0.02, -0.01, 0\}$. Values of γ imply an assumption about the percentage, d, of the effect of X on group-level differences in enrollment which operate through Z.

d, of the effect of X on group-level differences in enrollment which operate through Z. This exercise shows that unless all ($d = 100\%$) of the effect of socioeconomic status in explaining group-level differences in special education receipt operates through disability, the proxy-based model will understate the level of selection-contributing inequality. We emphasizes that $d = 100\%$ implies a very unlikely scenario since, as highlighted by the case of parental involvement discussed above, sociologists and educators are well-aware of pathways through which socioeconomic status might shape special education receipt even net of disability.

Assessment of empirical exercise

This exercise reveals serious limitations in using test scores and, to a lesser extent, socioeconomic status to control for disability status when measuring racial disproportionality in special education. For test scores to fully capture disability differences, they would need to show an implausibly large achievement gap between disabled and non-disabled students - over 4 standard deviations, or near 6.5 times larger than empirically observed. Alternatively, given an achievement gap between disabled and non-disabled students, the implied disability rate gap between Black and White students is impossibly large, at 117%. Even if test scores captured just 25% of true disability differences—meaning that the Black-White gap in special education receipt is nearly zero—the implied disability achievement gap would still be nearly twice what research suggests, or the implied differences in disability rates would be about nine times larger than they are currently. These calculations mean that existing proxy-based estimates understate the extent to which special education identification processes contribute to racial inequality. More to the point, it is almost certainly the case that the consensus of *under*-representation of Black students in special education is exaggerated. Instead, our analysis suggests that commonly used statistical controls may systematically mask real disparities in how students are identified for services.

7 Discussion

In this paper, we reflected on the methodological practice of using observational data to determine whether a selection process is *contributing* to inequality as opposed to simply *reflecting* inequalities that have emerged prior to the selection process itself. To clearly articulate the quantities of interest we propose new terminology detailing that the task at hand is to define whether a selection process produces *selection-contributing* or *selection-reflecting* inequality.

Identifying whether a selection process produces *selection-contributing* (as opposed to, or in addition to, *selection-reflecting*) inequality is a question of great sociological and policy relevance and concerns selection processes spanning different domains, including hiring decisions, within-firm promotions decisions, mortgage lending decisions, college/higher education admissions, a selection into various educational programs (e.g., Advanced Placement courses, gifted and talented programs and special education services).

Given the limited possibility of field experiments, much of this literature leverages observational data to assess selection-contributing inequality. This paper provides two central contributions to this line of quantitative inquiry.

First, we provide a conceptual framework to clearly define the quantity of interest. We suggest that to estimate selection-contributing inequality across groups, the researcher should compare the selection outcomes of different-group individuals who (at the time of the selection decision) similarly meet the criteria governing the selection process. This frameworks provides two important insights. First, statistical attempts to assess the existence of selection-contributing inequality cannot avoid the normative (and challenging) task of clearly defining what is (or should be) the criteria governing the selection process of interest. Second, a central challenge for researchers is to operationalize the latent concept of interest, the extent to which individuals meet the selection criteria. Implicitly, therefore, the goal of introducing a series of variables as statistical controls is to try to proxy for this latent concept.

Second, by providing this framework, we illustrate that traditional approaches have overlooked an important statistical issue, which we call the correlated proxy problem. Traditionally, studies rely on a kitchen-sink approach that compares the selection outcomes between different social groups net of a series of controls. This practice is informed by the understanding that the statistical issue preventing this kitchen-sink approaches from accurately capturing selection-contributing inequality is the issue of omitted-variable bias, i.e., the failure to control for group-level criteria-related differences. We show that this is not the case. If the chosen control (or proxy, to use the language proposed here) is correlated with other variables in the model in ways that are not fully explained by criteria-related differences (what we call the correlated proxy problem), then estimates can yield inaccurate estimates of selection-contributing inequality *even when omitted-variable bias is not present.*

Throughout the text, we have illustrated (and empirically demonstrated) the correlated proxy problem in the context of racial disproportionality in receipt of special education services in US schools. A central reason for choosing this example is that IDEA provides a clear definition of the selection criteria for special education, particularly in the context of specific leaning disabilities, clearly stipulating that educator professionals rule out some causes of lower academic performance, called "exclusionary factors", such as ex ante differences in educational opportunities. Then, by definition, if the chosen proxy explains racial differences in program enrollment through such exclusionary factors, a correlated proxy problem can arise. Special education services are, therefore, uniquely well suited to make it clear that the chosen proxy should capture criteria-related group-level differences while not capturing criteria-unrelated ones. And, as we have emphasized throughout this paper, while educator professionals have been charged with parsing the dual signals contained in a test score, researchers have not.

The correlated proxy problem extends beyond special education. In hiring decisions, researchers often control for educational credentials as a proxy for job qualification, but these credentials may capture group differences in interview preparation or networking skills unrelated to actual job capability. In child protective services, poverty indicators are commonly used to proxy for child neglect/abuse risk, but these may reflect neighborhood characteristics that influence official decisions through pathways unrelated to actual child welfare concerns. Similarly, in college admissions, essay quality might proxy for college readiness, but could instead capture access to tutoring resources rather than true academic preparation. In each case, proxy variables may explain group differences through pathways unrelated to the selection-relevant criteria, potentially masking true selection-contributing inequality. In Appendix E we provide further details for these applications to the correlated proxy problem.

Implications for future research

Our analysis demonstrates fundamental limitations in using regression-based methods to estimate selection-contributing inequality. The correlated proxy problem, like omitted-variable bias, can systematically bias estimates in ways that mislead policy and practice. This is particularly concerning given the widespread use of kitchen-sink regression approaches to assess selection-contributing inequality across many domains.

Two key methodological improvements emerge from our analysis. First, researchers must explicitly define the normative selection criteria governing the selection process before choosing control variables. This clarifies what characteristics are relevant for comparing groups and helps identify appropriate proxy measures. Second, researchers need to consider both omitted variable bias and the correlated proxy problem when selecting controls and interpreting results. We demonstrate how directed acyclic graphs and sensitivity analyses can help assess these dual sources of bias, though they cannot fully resolve the fundamental challenge of measuring selection-contributing inequality with observational data.

Recent methodological innovations such as the threshold test (Pierson et al. 2018; Simoiu et al. 2017), risk-adjusted regression (Grossman et al. 2023, 2024a; Jung et al. 2024; Souto-Maior and Shroff 2024), and variations of well-known benchmark and outcome tests (Gaebler et al. 2022; Gaebler and Goel 2024) offer promising alternatives to traditional regression approaches. Under specific circumstances, these approaches may better handle the inherent challenges of measuring selection-contributing inequality using observational data.

For special education specifically, our analysis suggests that existing research has not adequately measured whether Black and White students who are comparable *in terms of selection-relevant characteristics*—i.e., disability—have different likelihoods of receiving services. We demonstrate that the current practice of using test scores and socioeconomic status as proxies for disability or clinical need for services does not produce plausible estimates of inequality created by disability identification procedures. Research claiming Black students are under-represented are, at best, less definitive than previously suggested. Therefore, current debates about whether Black children are over- or under-represented in special education, relative to comparable White children, remain unresolved. This uncertainty leaves policymakers and education leaders without clear evidence on the magnitude of selection-contributing inequality for children receiving special education services, but our findings suggest that policy should not be altered to encourage greater identification of Black students with disabilities. Further research is crucial to understanding and addressing inequality in special education.

To better understand the role special education selection mechanisms have in contributing (or not) to racial inequality, new measures of disability status would be especially useful. As one example, such a measure may be constructed by more closely tying existing educator professional practices—those that require referral teams to determine whether exclusionary factors are causes of a student's lower academic performance—to secondary data collection. These assessments may be transferable to quantitative data and serve as representations of the desired latent variable. Other empirical approaches that do not rely on observational data, such as audit studies, will continue to have value given the challenges summarized here.

References

Akbaş, Merve, Dan Ariely, and Sevgi Yuksel. 2019. "When is inequality fair? An experiment on the effect of procedural justice and agency." *Journal of Economic Behavior & Organization* 161:114–127.

Angrist, Joshua D and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arcidiacono, Peter, Josh Kinsler, and Tyler Ransom. 2022. "Asian American discrimination in Harvard admissions." *European Economic Review* 144:104079.

Artiles, Alfredo J. 2013. "Untangling the racialization of disabilities: An intersectionality critique across disability models." *Du Bois Review: Social Science Research on Race* 10:329–347.

Artiles, Alfredo J, Beth Harry, Daniel J Reschly, and Philip C Chinn. 2002. "Over-identification of students of color in special education: A critical overview." *Multicultural perspectives* 4:3–10.

Ayres, Ian. 2005. "Three tests for measuring unjustified disparate impacts in organ transplantation: The problem of 'included variable' bias." *Perspectives in biology and medicine* 48:68–S87.

Ayres, Ian. 2010. "Testing for discrimination and the problem of 'included variable bias'." Technical report, Yale Law School (Working paper).

Baglivio, Michael T, Kevin T Wolff, Alex R Piquero, Mark A Greenwald, and Nathan Epps. 2017. "Racial/ethnic disproportionality in psychiatric diagnoses and treatment in a sample of serious juvenile offenders." *Journal of youth and adolescence* 46:1424–1451.

Balsa, Ana I, Thomas G McGuire, and Lisa S Meredith. 2005. "Testing for statistical discrimination in health care." *Health services research* 40:227–252.

Barnum, Matt. 2018. "DeVos presses pause on special education rule, highlighting ongoing discrimination debate." *Chalkbeat* .

Baron, Reuben M and David A Kenny. 1986. "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *Journal of personality and social psychology* 51:1173.

Barrett, Nathan, Andrew McEachin, Jonathan N Mills, and Jon Valant. 2021. "Disparities and discrimination in student discipline by race and family income." *Journal of Human Resources* 56:711–748.

Blau, Francine D and Jed DeVaro. 2007. "New evidence on gender differences in promotion rates: An empirical analysis of a sample of new hires." *Industrial Relations: A Journal of Economy and Society* 46:511–550.

Brand, Jennie E. 2015. “The far-reaching impact of job loss and unemployment.” *Annual review of sociology* 41:359–375.

Britto, Pia R, Stephen J Lye, Kerrie Proulx, Aisha K Yousafzai, Stephen G Matthews, Tyler Vaivada, Rafael Perez-Escamilla, Nirmala Rao, Patrick Ip, Lia CH Fernald, et al. 2017. *Nurturing care: promoting early childhood development*, volume 389. Elsevier.

Burke, Lindsey M. 2024. “Department of Education.” In *Mandate for Leadership 2025: The Conservative Promise*, edited by Paul Dans and Steven Groves, chapter 11, p. 319. The Heritage Foundation.

Calarco, Jessica McCrory. 2018. *Negotiating opportunities: How the middle class secures advantages in school*. Oxford, UK: Oxford University Press.

Christopher, Karen. 1996. “Explaining the recent employment gap between black and white women.” *Sociological Focus* 29:263–280.

Cichy, Kelly E, Robert S Stawski, and David M Almeida. 2012. “Racial differences in exposure and reactivity to daily family stressors.” *Journal of Marriage and Family* 74:572–586.

Cinelli, Carlos, Andrew Forney, and Judea Pearl. 2024. “A crash course in good and bad controls.” *Sociological Methods & Research* 53:1071–1104.

Conger, Dylan, Mark C Long, and Patrice Iatarola. 2009. “Explaining race, poverty, and gender disparities in advanced course-taking.” *Journal of Policy Analysis and Management* 28:555–576.

Cowhy, Jennifer R, Quinn Mulroy, and Tabitha Bonilla. 2024. “Reconceptualizing Parents as Policy Agents Within Special Education.” *Educational Researcher* p. 0013189X241238679.

Cruz, Rebecca A and Janelle E Rodl. 2018. “An integrative synthesis of literature on disproportionality in special education.” *The Journal of Special Education* 52:50–63.

Ding, Peng and Luke W Miratrix. 2015. “To adjust or not to adjust? Sensitivity analysis of M-bias and butterfly-bias.” *Journal of Causal Inference* 3:41–57.

DiPrete, Thomas A and Whitman T Soule. 1988. “Gender and promotion in segmented job ladder systems.” *American Sociological Review* pp. 26–40.

Domina, Thurston, Andrew Penner, and Emily Penner. 2017. “Categorical inequality: Schools as sorting machines.” *Annual review of sociology* 43:311–330.

Donovan, M Suzanne and Christopher T Cross. 2002. *Minority students in special and gifted education*. National Academies Press.

Dunn, Lloyd M. 1968. “Special education for the mildly retarded—Is much of it justifiable?” *Exceptional children* 35:5–22.

Edwards, Linsey. 2016. "Homogeneity and inequality: School discipline inequality and the role of racial composition." *Social Forces* 95:55–76.

Elwert, Felix and Christopher Winship. 2014. "Endogenous selection bias: The problem of conditioning on a collider variable." *Annual review of sociology* 40:31–53.

Espenshade, Thomas J, Chang Y Chung, and Joan L Walling. 2004. "Admission preferences for minority students, athletes, and legacies at elite universities." *Social Science Quarterly* 85:1422–1446.

Farber, Henry S and Robert G Valletta. 2015. "Do extended unemployment benefits lengthen unemployment spells?: Evidence from recent cycles in the US labor market." *Journal of Human Resources* 50:873–909.

Farkas, George and Keven Vicknair. 1996. "Appropriate tests of racial wage discrimination require controls for cognitive skill: Comment on Cancio, Evans, and Maume." *American Sociological Review* 61:557–560.

Feely, Megan and Emily Adlin Bosk. 2021. "That which is essential has been made invisible: The need to bring a structural risk perspective to reduce racial disproportionality in child welfare." *Race and social problems* 13:49–62.

Fernandez, Roberto M and M Lourdes Sosa. 2005. "Gendering the job: Networks and recruitment at a call center." *American Journal of Sociology* 111:859–904.

Fish, Rachel Elizabeth. 2017. "The racialized construction of exceptionality: Experimental evidence of race/ethnicity effects on teachers' interventions." *Social Science Research* 62:317–334.

Fish, Rachel Elizabeth. 2019. "Standing out and sorting in: Exploring the role of racial composition in racial disparities in special education." *American Educational Research Journal* 56:2573–2608.

Fish, Rachel Elizabeth, Kenneth Shores, and Joao Souto-Maior. 2024. "A Critical Appraisal of the Evidence on Racial Disproportionality in Special Education." *OSF preprint* .

Fletcher, Jack M, G Reid Lyon, Lynn S Fuchs, and Marcia A Barnes. 2018. *Learning disabilities: From identification to intervention*. Guilford Publications.

Fong, Kelley. 2020. "Getting eyes in the home: Child protective services investigations and state surveillance of family life." *American Sociological Review* 85:610–638.

Font, Sarah A, Lawrence M Berger, and Kristen S Slack. 2012. "Examining racial disproportionality in child protective services case decisions." *Children and youth services review* 34:2188–2200.

Fuller, Joseph, Christina Langer, and Matt Sigelman. 2022. "Skills-based hiring is on the rise." *Harvard Business Review* 11:1–6.

Gaddis, S Michael. 2018. *An introduction to audit studies in the social sciences*. Springer.

Gaebler, Johann, William Cai, Guillaume Basse, Ravi Shroff, Sharad Goel, and Jennifer Hill. 2022. "A causal framework for observational studies of discrimination." *Statistics and public policy* 9:26–48.

Gaebler, Johann D and Sharad Goel. 2024. "A Simple, Statistically Robust Test of Discrimination." *arXiv preprint arXiv:2407.06539* .

Gelman, Andrew, Jeffrey Fagan, and Alex Kiss. 2007. "An analysis of the New York City police department's "stop-and-frisk" policy in the context of claims of racial bias." *Journal of the American statistical association* 102:813–823.

Gilmour, Allison F, Douglas Fuchs, and Joseph H Wehby. 2019. "Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities." *Exceptional Children* 85:329–346.

Goel, Sharad, Justin M Rao, Ravi Shroff, et al. 2016. "Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy." *The Annals of Applied Statistics* 10:365–394.

Gordon, Nora. 2017. "Race, poverty, and interpreting overrepresentation in special education." https://www.brookings.edu/articles/race-poverty-and-interpreting-overrepresentation-in-special-education/. Accessed on December 13, 2024.

Gotham, Kevin Fox. 1998. "Race, mortgage lending and loan rejections in a US city." *Sociological Focus* 31:391–405.

Grissom, Jason A and Christopher Redding. 2015. "Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs." *Aera Open* 2:2332858415622175.

Grossman, Joshua, Julian Nyarko, and Sharad Goel. 2023. "Racial bias as a multi-stage, multi-actor problem: An analysis of pretrial detention." *Journal of Empirical Legal Studies* 20:86–133.

Grossman, Joshua, Julian Nyarko, and Sharad Goel. 2024a. "Reconciling Legal and Empirical Conceptions of Disparate Impact: An Analysis of Police Stops Across California." *Journal of Law and Empirical Analysis* 1:2755323X241243168.

Grossman, Joshua, Sabina Tomkins, Lindsay Page, and Sharad Goel. 2024b. "The disparate impacts of college admissions policies on Asian American applicants." *Scientific Reports* 14:4449.

Grusky, David. 1994. *Social stratification, class, race, and gender in sociological perspective*. Routledge.

Harry, Beth and Janette Klingner. 2014. *Why are so many minority students in special education?* Teachers College Press.

Hemmerechts, Kenneth, Orhan Agirdag, and Dimokritos Kavadias. 2017. "The relationship between parental literacy involvement, socio-economic status and reading literacy." *Educational Review* 69:85–101.

Hibel, Jacob, Susan Faircloth, and George Farkas. 2008. "Unpacking the placement of American Indian and Alaska Native students in special education programs and services in the early grades: School readiness as a predictive variable." *Harvard Educational Review* 78:498–528.

Hibel, Jacob, George Farkas, and Paul L Morgan. 2010. "Who is placed into special education?" *Sociology of education* 83:312–332.

Hout, Michael and Thomas A DiPrete. 2006. "What we have learned: RC28's contributions to knowledge about social stratification." *Research in social stratification and mobility* 24:1–20.

Huang, Francis L. 2020. "Prior problem behaviors do not account for the racial suspension gap." *Educational Researcher* 49:493–502.

IDEA. 2004. "Individuals with Disabilities Education Act of 2004." https://sites.ed.gov/idea/.

IDEA-Sec.300.309. 2006. "Individuals with Disabilities Education Act of 2004. Sec. 300.309: Determining the existence of a specific learning disability." FR 46753, Aug. 14, 2006, as amended at 82 FR 31912, July 11, 2017. https://sites.ed.gov/idea/regs/b/d/300.309.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A general approach to causal mediation analysis." *Psychological methods* 15:309.

Imhof, Andrea, Sihong Liu, Lisa Schlueter, Tiffany Phu, Sarah Watamura, and Philip Fisher. 2023. "Improving children's expressive language and auditory comprehension through responsive caregiving: evidence from a randomized controlled trial of a strength-based video-coaching intervention." *Prevention Science* 24:84–93.

Irizarry, Yasmiyn. 2021. "On Track or Derailed? Race, Advanced Math, and the Transition to High School." *Socius: Sociological Research for a Dynamic World* 7:1–21.

Jayakumar, Uma Mazyck and Scott E Page. 2021. "Cultural capital and opportunities for exceptionalism: Bias in university admissions." *The Journal of Higher Education* 92:1109–1139.

Jung, Jongbin, Sam Corbett-Davies, Ravi Shroff, and Sharad Goel. 2024. "Omitted and included variable bias in tests for disparate impact." *arXiv preprint arXiv:1809.05651* .

Kelly, Sean. 2009. "The black-white gap in mathematics course taking." *Sociology of Education* 82:47–69.

Kincaid, Aleksis P and Amanda L Sullivan. 2017. "Parsing the relations of race and socioeconomic status in special education disproportionality." *Remedial and Special Education* 38:159–170.

Lanphear, Bruce P, Michael Weitzman, and Shirley Eberly. 1996. "Racial differences in Urban children's environmental exposures to lead." *American journal of public health* 86:1460–1463.

Lareau, Annette. 2011. *Unequal childhoods: class, race, and family life*. Berkeley, CA: Univ of California Press.

Legewie, Joscha. 2016. "Racial profiling and use of force in police stops: How local events trigger periods of increased discrimination." *American journal of sociology* 122:379–424.

Lewis, Amanda E and John B Diamond. 2015. *Despite the best intentions: How racial inequality thrives in good schools*. Oxford University Press.

Lewis-McCoy, RL'Heureux. 2014. *Inequality in the promised land: Race, resources, and suburban schooling*. Stanford University Press.

Lind, E Allan and Tom R Tyler. 2013. *The social psychology of procedural justice*. Springer Science & Business Media.

Long, Daniel A, D Betsy McCoach, Del Siegle, Carolyn M Callahan, and E Jean Gubbins. 2023. "Inequality at the starting line: Underrepresentation in gifted identification and disparities in early achievement." *AERA Open* 9:23328584231171535.

Lucas, Samuel. 2009. *Theorizing discrimination in an era of contested prejudice: Discrimination in the United States*. Temple University Press.

Lucas, Samuel R, Santiago J Molina, and John M Towey. 2020. "Race/Ethnicity over Fifty Years of Structural Differentiation in K–12 Schooling: Period-Specific and Life-Course Perspectives." *Annual Review of Sociology* 46:355–378.

Lundberg, Ian, Rebecca Johnson, and Brandon M Stewart. 2021. "What is your estimand? Defining the target quantity connects statistical evidence to theory." *American Sociological Review* 86:532–565.

MacCoun, Robert J. 2005. "Voice, control, and belonging: The double-edged sword of procedural fairness." *Annu. Rev. Law Soc. Sci.* 1:171–201.

Maroto, Michelle, David Pettinicchio, and Andrew C Patterson. 2019. "Hierarchies of categorical disadvantage: Economic insecurity at the intersection of disability, gender, and race." *Gender & Society* 33:64–93.

Massey, Douglas S. 2007. *Categorically unequal: The American stratification system*. Russell Sage Foundation.

Mayes, Keith A. 2023. *The unteachables: Disability rights and the invention of Black special education*. U of Minnesota Press.

Mendive, Susana, María Rosa Lissi, Roger Bakeman, and Adriana Reyes. 2017. "Beyond mother education: Maternal practices as predictors of early literacy development in Chilean children from low-SES households." *Early Education and Development* 28:167–181.

Mickelson, Roslyn Arlin. 2001. "Subverting Swann: First-and second-generation segregation in the Charlotte-Mecklenburg schools." *American Educational Research Journal* 38:215–252.

Miller, Dale T. 2001. "Disrespect and the experience of injustice." *Annual review of psychology* 52:527–553.

Mize, Trenton D. 2019. "Best practices for estimating, interpreting, and presenting nonlinear interaction effects." *Sociological Science* 6:81–117.

Moll, Kristina, Sarah Kunze, Nina Neuhoff, Jennifer Bruder, and Gerd Schulte-Körne. 2014. "Specific learning disorder: Prevalence and gender differences." *PLoS one* 9:e103537.

Mood, Carina. 2010. "Logistic regression: Why we cannot do what we think we can do, and what we can do about it." *European Sociological Review* 26:67–82.

Morgan, Paul L. 2021. "Unmeasured confounding and racial or ethnic disparities in disability identification." *Educational Evaluation and Policy Analysis* 43:351–361.

Morgan, Paul L and George Farkas. 2015. "Is Special Education Racist?" *New York Times* June 24, 2015. https://www.nytimes.com/2015/06/24/opinion/is-special-education-racist.html.

Morgan, Paul L, George Farkas, Michael Cook, Natasha M Strassfeld, Marianne M Hillemeier, Wik Hung Pun, and Deborah L Schussler. 2017. "Are Black children disproportionately overrepresented in special education? A best-evidence synthesis." *Exceptional Children* 83:181–198.

Morgan, Paul L, George Farkas, Marianne M Hillemeier, and Steve Maczuga. 2012. "Are minority children disproportionately represented in early intervention and early childhood special education?" *Educational Researcher* 41:339–351.

Morgan, Paul L, George Farkas, Marianne M Hillemeier, Richard Mattison, Steve Maczuga, Hui Li, and Michael Cook. 2015. "Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions." *Educational Researcher* 44:278–292.

Morgan, Paul L, Adrienne D Woods, Yangyang Wang, Marianne M Hillemeier, George Farkas, and Cynthia Mitchell. 2020. "Are schools in the US South using special education to segregate students by race?" *Exceptional Children* 86:255–275.

Muñiz, Julissa O. 2021. "Exclusionary discipline policies, school-police partnerships, surveillance technologies and disproportionality: A review of the school to prison pipeline literature." *The Urban Review* 53:735–760.

Pager, Devah and Hana Shepherd. 2008. "The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets." *Annu. Rev. Sociol* 34:181–209.

Pearl, Judea. 2000. *Models, reasoning and inference*. Cambridge University Press.

Pearl, Judea. 2014. "Interpretation and identification of causal mediation." *Psychological methods* 19:459.

Pearman, Francis A and Ebony O McGee. 2022. "Anti-Blackness and racial disproportionality in gifted education." *Exceptional Children* 88:359–380.

Pedulla, David. 2020. *Making the cut: Hiring decisions, bias, and the consequences of nonstandard, mismatched, and precarious employment*. Princeton University Press.

Peters, Scott J, Marcia Gentry, Gilman W Whiting, and Matthew T McBee. 2019. "Who gets served in gifted education? Demographic representation and a call for action." *Gifted Child Quarterly* 63:273–287.

Petersen, Trond and Ishak Saporta. 2004. "The opportunity structure for discrimination." *American Journal of Sociology* 109:852–901.

Petersen, Trond, Ishak Saporta, and Marc-David L Seidel. 2005. "Getting hired: Sex and race." *Industrial Relations: A Journal of Economy and Society* 44:416–443.

Petersen, Trond and Thea Togstad. 2006. "Getting the offer: Sex discrimination in hiring." *Research in Social Stratification and Mobility* 24:239–257.

Pierson, Emma, Sam Corbett-Davies, and Sharad Goel. 2018. "Fast threshold tests for detecting discrimination." *International conference on artificial intelligence and statistics* pp. 96–105.

Putnam-Hornstein, Emily, Barbara Needell, Bryn King, and Michelle Johnson-Motoyama. 2013. "Racial and ethnic disparities: A population-based examination of risk factors for involvement with child protective services." *Child abuse & neglect* 37:33–46.

Quillian, Lincoln, John J Lee, and Brandon Honoré. 2020. "Racial discrimination in the US housing and mortgage lending markets: a quantitative review of trends, 1976–2016." *Race and Social Problems* 12:13–28.

Quillian, Lincoln and Arnfinn H Midtbøen. 2021. "Comparative perspectives on racial discrimination in hiring: The rise of field experiments." *Annual Review of Sociology* 47:391–415.

Quinn, David M. 2015. "Kindergarten Black–White test score gaps: Re-examining the roles of socioeconomic status and school quality with new data." *Sociology of Education* 88:120–139.

Ray, Rashawn, Connor Powelson, Genesis Fuentes, and Long Doan. 2024. "The Sociology of Police Behavior." *Annual Review of Sociology* 50.

Reardon, Sean F, Demetra Kalogrides, and Kenneth Shores. 2019a. "The geography of racial/ethnic test score gaps." *American Journal of Sociology* 124:1164–1221.

Reardon, Sean F, Ericka S Weathers, Erin M Fahle, Heewon Jang, and Demetra Kalogrides. 2019b. "Is separate still unequal? New evidence on school segregation and racial academic achievement gaps." *American Sociological Review* pp. 973–1010.

Riegle-Crumb, Catherine, Barbara King, and Yasmiyn Irizarry. 2019. "Does STEM stand out? Examining racial/ethnic gaps in persistence across postsecondary fields." *Educational Researcher* 48:133–144.

Rivaux, Stephanie L, Joyce James, Kim Wittenstrom, Donald Baumann, Janess Sheets, Judith Henry, and Victoria Jeffries. 2008. "The intersection of race, poverty, and risk." *Child welfare* 87:151–168.

Rivera, Lauren A and András Tilcsik. 2019. "Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation." *American Sociological Review* 84:248–274.

Rosenbaum, Paul R. 1984. "The consequences of adjustment for a concomitant variable that has been affected by the treatment." *Journal of the Royal Statistical Society Series A: Statistics in Society* 147:656–666.

Royce, Edward. 2022. *Poverty and power: The problem of structural inequality*. Rowman & Littlefield.

Sampson, Robert J. 2008. "Moving to inequality: Neighborhood effects and experiments meet social structure." *American journal of sociology* 114:189–231.

Sampson, Robert J and Alix S Winter. 2016. "The racial ecology of lead poisoning: Toxic inequality in Chicago neighborhoods, 1995-2013." *Du Bois Review: Social Science Research on Race* 13:261–283.

Shakespeare, Tom. 2013. *Disability rights and wrongs revisited*. Routledge.

Shifrer, Dara. 2018. "Clarifying the social roots of the disproportionate classification of racial minorities and males with learning disabilities." *The Sociological Quarterly* 59:384–406.

Shifrer, Dara and Angela Frederick. 2019. "Disability at the intersections." *Sociology Compass* 13:e12733.

Shores, Kenneth, Ha Eun Kim, and Mela Still. 2020. "Categorical inequality in Black and White: Linking disproportionality across multiple educational outcomes." *American Educational Research Journal* 57:2089–2131.

Simoiu, Camelia, Sam Corbett-Davies, and Sharad Goel. 2017. "The problem of infra-marginality in outcome tests for discrimination." *The Annals of Applied Statistics* .

Skaggs, Sheryl and Jennifer Bridges. 2013. "Race and sex discrimination in the employment process." *Sociology Compass* 7:404–415.

Skiba, Russell J, Alfredo J Artiles, Elizabeth B Kozleski, Daniel J Losen, and Elizabeth G Harry. 2016a. "Risks and consequences of oversimplifying educational inequities: A response to Morgan et al.(2015)." *Educational Researcher* 45:221–225.

Skiba, Russell J, Kavitha Mediratta, and M Karega Rausch. 2016b. *Inequality in school discipline: Research and practice to reduce disparities*. Springer.

Skiba, Russell J, Lori Poloni-Staudinger, Ada B Simmons, L Renae Feggins-Azziz, and Choong-Geun Chung. 2005. "Unproven links: Can poverty explain ethnic disproportionality in special education?" *The Journal of Special Education* 39:130–144.

Souto-Maior, João M and Ravi Shroff. 2024. "Differences in Academic Preparedness Do Not Fully Explain Black–White Enrollment Disparities in Advanced High School Coursework." *Sociological Science* 11:138–163.

Spencer, Karen Lutfey and Matthew Grace. 2016. "Social foundations of health care inequality and treatment bias." *Annual Review of Sociology* 42:101–120.

Sullivan, Amanda L and Aydin Bal. 2013. "Disproportionality in special education: Effects of individual and school variables on disability risk." *Exceptional Children* 79:475–494.

Tilly, Charles. 1998. *Durable inequality*. University of California Press.

Tyson, Karolyn. 2011. *Integration interrupted: Tracking, Black students, and acting White after Brown*. New York, NY: Oxford University Press.

U.S. Department of Education. 2019. "IDEA Part B Data Collections." EDFacts Data Warehouse. Includes "IDEA Part B Discipline Collection" and "IDEA Part B Child Count and Educational Environments Collection" for 2018-19. Available at: https://www2.ed.gov/programs/osepidea/618-data/state-level-data-files/.

Warikoo, Natasha. 2022. *Is Affirmative Action Fair?: The Myth of Equity in College Admissions*. John Wiley & Sons.

Appendix A Derivation of the correlated proxy problem

General strategy

For the ideal model, we estimate a regression model of the form $Y = \beta_0 + \beta_1^{raw} A + \epsilon$. Estimating it through OLS ($\mathrm{E}(\epsilon) = 0$), we can write $\hat{\beta}_1^{raw}$ as:

$$\hat{\beta}_1^{raw} = \frac{\mathrm{Cov}(Y, G)}{\mathrm{Var}(G)} \tag{13}$$

For the proxy-based model, we estimate a regression of the form $Y = \beta_0 + \beta_1^{proxy} G + \beta_2 X + \epsilon$. Estimating it through OLS ($\epsilon \perp\!\!\!\perp (G, X)$ and $\mathrm{E}(\epsilon) = 0$), we can solve for $(\hat{\beta}_1^{proxy}, \hat{\beta}_2)$, using the following two moment conditions (Ding and Miratrix 2015):

$$\begin{cases} \mathrm{Cov}(Y, G) = \hat{\beta}_1^{proxy} \mathrm{Var}(G) + \hat{\beta}_2 \mathrm{Cov}(G, X) \\ \mathrm{Cov}(Y, X) = \hat{\beta}_1^{proxy} \mathrm{Cov}(G, X) + \hat{\beta}_2 \mathrm{Var}(X) \end{cases} \tag{14}$$

From these two moment conditions, we can write $\hat{\beta}_1^{proxy}$ as:

$$\hat{\beta}_1^{proxy} = \frac{\mathrm{Cov}(Y, G)\mathrm{Var}(B) - \mathrm{Cov}(Y, X)\mathrm{Cov}(G, X)}{\mathrm{Var}(G)\mathrm{Var}(X) - \mathrm{Cov}(G, X)^2} \tag{15}$$

For the ideal model, we estimate a regression of the form $Y = \beta_0 + \beta_1^{ideal} G + \beta_2 Z + \epsilon$. Then, analogously to 15, we can write the OLS estimator for $\hat{\beta}^{proxy}$ as:

$$\hat{\beta}_1^{ideal} = \frac{\mathrm{Cov}(Y, G)\mathrm{Var}(B) - \mathrm{Cov}(Y, Z)\mathrm{Cov}(G, Z)}{\mathrm{Var}(G)\mathrm{Var}(Z) - \mathrm{Cov}(G, Z)^2} \tag{16}$$

Then, to solve for $\hat{\beta}_1^{raw}$, $\hat{\beta}_1^{proxy}$ and $\hat{\beta}_1^{ideal}$ for a given data-generating process, we need to derive the values of $\mathrm{Var}(G)$, $\mathrm{Var}(X)$, $\mathrm{Var}(Z)$, $\mathrm{Cov}(Y, G)$, $\mathrm{Cov}(Y, X)$, $\mathrm{Cov}(Y, Z)$, $\mathrm{Cov}(G, X)$ and $\mathrm{Cov}(G, Z)$. We can do so by, first, writing endogenous variables in terms of the exogenous variables in the system and, second, by using the properties of variances and covariances — i.e., for random variables X and Y and constants a and b, we have that $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$ and that $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.

We apply this approach to solve for $\hat{\beta}_1^{raw}$, $\hat{\beta}_1^{proxy}$ and $\hat{\beta}_1^{ideal}$ given the data-generating processes defined in the main text, LSEM in Eq. 1 (for a type-A proxy) and LSEM in Eq. 6 (for a type-B proxy).

Type-A proxy

First, we write variables Z, X and Y only in terms of the exogenous variables in the system, G. Variable Z is already defined only in terms of G:

$$Z = G\alpha_1 + \epsilon_1. \tag{17}$$

To do the same for X, we replace $Z = G\alpha_1 + \epsilon_1$ into the equation for X, writing it as:

$$X = G\gamma + \alpha_2(G\alpha_1 + \epsilon_1) + \epsilon_2. \tag{18}$$

Then, we use Eq.18 to replace X in the equation for Y, writing Y as:

$$Y = G\alpha_4 + \alpha_3(G\gamma + \alpha_2(G\alpha_1 + \epsilon_1) + \epsilon_2) + \epsilon_3. \tag{19}$$

Then, we solve for variances and covariances of interest. To illustrate, consider $\text{Cov}(Y, Z)$, which can be written as the product of Eqs.17 and 19. Using the fact that $\text{Var}(G) = p(1-p)$ and the properties of variances and covariances, we can $\text{Cov}(Y, Z)$ as:

$$\text{Cov}(Y, Z) = \alpha_1^2\alpha_2\alpha_3 p(1-p) + \alpha_1\alpha_3\gamma p(1-p) + \alpha_1\alpha_4 p(1-p) + \alpha_2\alpha_3 \tag{20}$$

We do the same for all the variances and covariances of interest. Replacing all solutions into Eqs. 13, 15 and 16 and simplifying, we get:

$$\hat{\beta}_1^{raw} = \alpha_1\alpha_2\alpha_3 + \alpha_3\gamma + \alpha_4; \tag{21}$$

$$\hat{\beta}_1^{ideal} = \alpha_3\gamma + \alpha_4; \tag{22}$$

$$\hat{\beta}_1^{proxy} = \alpha_4. \tag{23}$$

Type-B proxy

As above, we, first, write variables Z, X and Y only in terms of the exogenous variables in the system, G. Variable X is already defined only in terms of G:

$$X = G\alpha_1 + \epsilon_1. \tag{24}$$

To do the same for Z, we replace $X = G\alpha_1 + \epsilon_1$ into the equation for Z, writing it as:

$$Z = \alpha_2(G\alpha_1 + \epsilon_1) + \epsilon_2. \tag{25}$$

Then, we replace Eqs. 24 and 25 into the equation for Y, writing it as:

$$Y = \alpha_3(\alpha_2(G\alpha_1 + \epsilon_1) + \epsilon_2) + G\alpha_4 + \gamma(G\alpha_1 + \epsilon_1) + \epsilon_3. \tag{26}$$

The, we solve for variances and covariances of interest. To illustrate, consider, again, $\text{Cov}(Y, Z)$, which can be written as the product of Eqs. 25 and 26. Using the fact that $\text{Var}(G) = p(1-p)$ and the properties of variances and covariances, we can $\text{Cov}(Y, Z)$ as:

$$\text{Cov}(Y, Z) = \alpha_1^2\alpha_2^2\alpha_3 p(1-p) + \alpha_1^2\alpha_2\gamma p(1-p) + \alpha_1\alpha_2\alpha_4 p(1-p) + \alpha_2^2\alpha_3 + \alpha_2\gamma + \alpha_3 \tag{27}$$

We do the same for all the variances and covariances of interest. Replacing all solutions into Eqs. 13, 15 and 16 and simplifying, we get:

$$\hat{\beta}_1^{raw} = \alpha_1\alpha_2\alpha_3 + \alpha_1\gamma + a_4; \tag{28}$$

$$\hat{\beta}_1^{ideal} = \frac{\alpha_1\gamma + \alpha_2^2\alpha_4 + \alpha_4}{\alpha_2^2 + 1}; \tag{29}$$

$$\hat{\beta}_1^{proxy} = \alpha_4. \tag{30}$$

Appendix B Supplemental figures

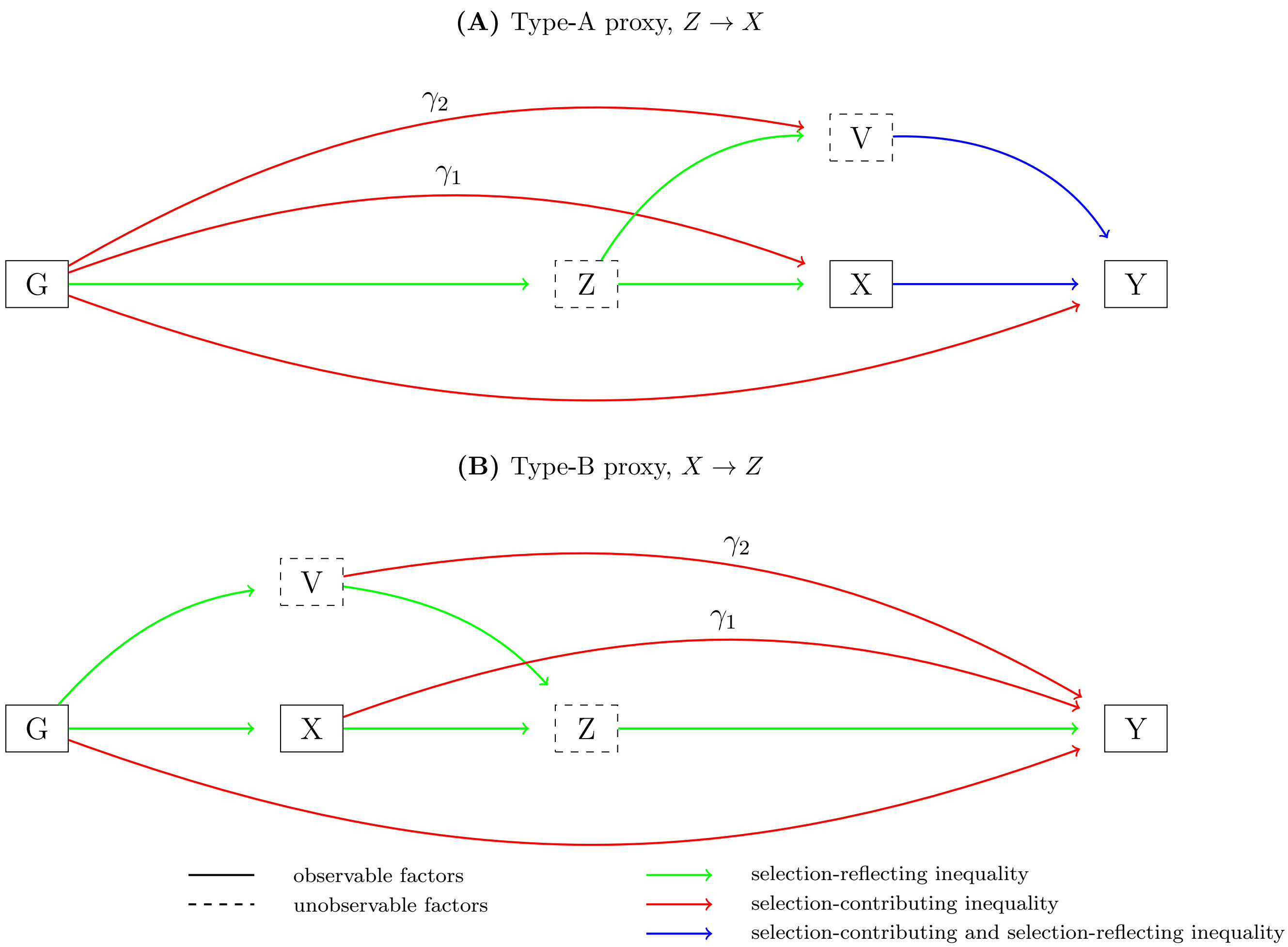


Figure B.1: The correlated proxy problem when two proxies are needed to meet the condition of no omitted-variable condition.

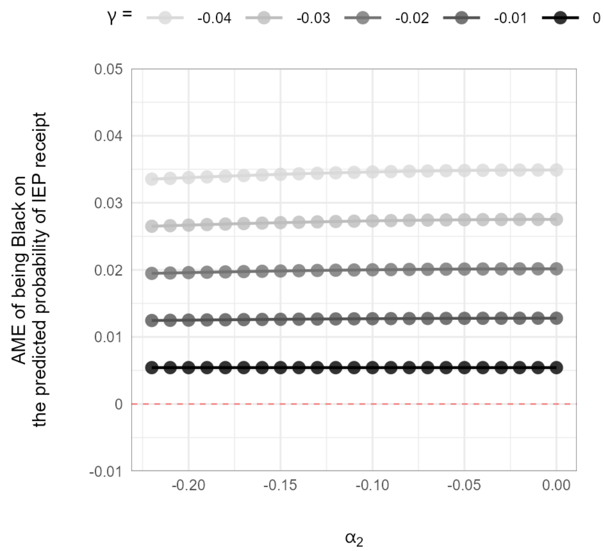


Figure B.2: Variations in the sensitivity estimate presented in Figure 8 in the main text for $\alpha_2 \in \{-0.22, -0.21, ..., -0.01, 0\}$. Parameter α_1 is known from data (= -0.74) and $\gamma \in \{-0.04, -0.03, -0.02, -0.01, 0\}$.

Appendix C Supplemental calculations for interpretation of Figure 7

The goal of this exercise is to identify a disability versus non-disability achievement gap that would be necessary to explain the totalizing influence of test scores for explaining racial disproportionality in special education. Defining terms, let t_{nd} be the average test score for students without disabilities; $t - d$ be the average test score for students with disabilities; and, finally, t_{nd-d} be the average difference in test score between students without and with disabilities ($t_{nd-d} = t_{nd} - t_d$). This number is not known given that the true disability rate is unobserved but can be inferred if we assign other values.

For example, following national-level demographics, we take as a population one composed of 15% Black students, 50% White students, and 35% others race groups. Then, assume an overall disability rate of 15% (this number is not known, given the latent variable problem but roughly matches current national estimates). Next, we allow for a Black-White disability gap that is 3:1, meaning that Black student disability rates are 3 times that of

White students (i.e., disability rate is 22.5% for Blacks and 7.5% for Whites). This, too, is not known, and greatly exceeds current estimates disproportionality, allowing for the possibility of true over-representation of Black students in the population. Then, following known test score inequality (e.g., Reardon et al. 2019a), let the Black-White test score gap for the overall population be 0.7 standard deviations. In sum, we have conservatively allowed for racial differences in disability rates to be much larger than are observed in practice. Even with this conservative allowance, as we will see, the implied necessary disability to non-disability achievement gap is implausibly large.

Given these initial conditions, we can solve for t_{nd-d} in terms of parameter d Figure 7, i.e., the percentage of the Black-White test score gap that is explained by disability, Z.

Given that the unconditional test score gap is 0.7, we can write the size of the gap to be explained as:

$$d0.7 = t_w - t_b \tag{31}$$

The average test score for Whites, t_w, can be written as:

$$t_w = 0.925(t_{nd}) + 0.075(t_{nd} - t_{nd-d}) \tag{32}$$

$$= t_{nd} - 0.075 t_{nd-d} \tag{33}$$

Similarly, the average test score for Blacks, t_b, can be written as:

$$t_b = 0.775(t_{nd}) + 0.225(t_{nd} - t_{nd-d}) \tag{34}$$

$$= t_{nd} - 0.225 t_{nd-d} \tag{35}$$

Given Eq. 31, have that:

$$t_b = t_w - d0.7. \tag{36}$$

Substituting Eqs. 33and 35 into the equation above:

$$t_{nd} - 0.225 t_{nd-d} = t_{nd} - 0.075 t_{nd-d} - d0.7, \tag{37}$$

which simplifies to:

$$d0.7 = 0.15t_{nd-d}, \tag{38}$$

and we find that:

$$t_{nd-d} = \frac{d0.7}{0.15} \tag{39}$$

Therefore, if $d = 100\%$, then the average difference in test scores between students without and with disabilities (t_{nd-d}) is about 4.67 standard deviations. Similarly, if $d = 25\%$, then the average difference in test score between students without and with disabilities (t_{nd-d}) should be about 1.17 standard deviations.

Manipulation of different values — for example, allowing the disability to non-disability achievement gap to be 0.62 SD (Gilmour et al. 2019) — allows us to calculate the implied disability rate gap between Black and White students. Here we now solve for the values 0.225 and 0.075 from above (i.e., defining the difference as d_{b-w} and solve for t_{nd-d}. If $d = 100\%$, then $d_{b-w} = 1.17$, an impossible number, and if $d = 25\%$, then $d_{b-w} = 0.282$, a value that is 9.4 times larger than current differences in the rates Black and White students receive special education services.

Appendix E Other examples of policy significance of the correlated proxy problem

Group-level disparities in hiring decisions

Broadly, we might say that hiring committees would like to hire the candidate who is best qualified for the job (the selection criteria). This latent concept is unobservable to the researcher. Then, to analyze selection-contributing inequality in the context of hiring practices, researchers often control for the candidate's educational background (e.g., whether the individual has a college degree). Such indicator might be a determinant of job qualification and, thus, might seem like a reasonable type-B proxy. However, it is not necessarily the case that individuals with different educational backgrounds have different levels of job qualification — in fact, firms increasingly perceive an ambiguity between educational credentials and job qualification and many are moving to skills-based hiring practices (Fuller et al. 2022). This imperfect correlation of the educational background with the latent concept of interest leaves room for educational background to explain group-level differences in hiring outcomes through pathways (e.g., the fact that college-goers might have received targeted training on how to appropriately navigate job interviews) that are unrelated to job qualification, path γ in Figure 5B's DAG. This means that while controlling for educational background might lead the model to explain group-level differences hiring outcomes, it is not necessarily the case that, in doing so, the model is explaining such group-level disparities in terms of differences in job qualification.

Group-level disparities in allocation to child protective services

According to federal guidelines, a child should be selected for child protective services based on wether they are experiencing child neglect or abuse (Fong 2020), the selection criteria. As in the example above, this latent concept is unobservable to the researcher. Common practice is to rely on poverty indicators as proxies for it. Poverty, by bringing about social and economic hardships, can be a determinant of child neglect/abuse. Thus, it can be seen as a reasonable type-B proxy. However, it is possible that other factors (e.g., neighborhood quality), which are correlated with poverty, shape influence child protective service official's decisions through pathways that are unrelated to the selection criteria, path γ in Figure 5A's DAG. Thus, controlling for poverty indicators might explain group-level differences in allocation to child protected services through pathways that are unrelated to the selection criteria.

Group-level disparities in college admissions

Consider that universities (arguably) should admit the students who are more likely to have academic success in college (perhaps by actually getting a college degree, what is often capture by the idea of college readiness (the selection criteria). Ideally, perhaps, one might think we should control, for instance, for essay content — a central factor in US college admissions. Intuitively, essay content can be seen as a manifestation of college readiness and, thus, it might be a reasonable type-A proxy. However, it is possible that other factors (e.g., parents' ability to afford a college tutor) shape essay context through pathways that are unrelated to college readiness, path γ in Figure 5A's DAG. This means that controlling for essay content might allow the model to explain group-level differences in admissions outcomes via group-level differences which do not exclusively capture college readiness.