View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarWorks @ Georgia State University

Georgia State University
ScholarWorks @ Georgia State University

Economics Dissertations

Department of Economics

Summer 8-1-2015

Evaluating Occupational Outcomes and Interventions in Schools

Julia Manzella

Follow this and additional works at: https://scholarworks.gsu.edu/econ_diss

Recommended Citation

Manzella, Julia, "Evaluating Occupational Outcomes and Interventions in Schools." Dissertation, Georgia State University, 2015.
https://scholarworks.gsu.edu/econ_diss/115

This Dissertation is brought to you for free and open access by the Department of Economics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Economics Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

ABSTRACT

EVALUATING OCCUPATIONAL OUTCOMES AND INTERVENTIONS IN SCHOOLS

BY

JULIA ANN MANZELLA

AUGUST 2015

Committee Chair: Dr. Barry T. Hirsch

Major Department: Economics

This dissertation consists of three distinct yet inter-related research papers in labor economics, each with relevance for public policy. The first chapter examines the role of wage differentials for caring work in explaining the gender wage gap. We find that both women and men face caring penalties that are small, about 2% for one standard deviation difference in caring. While women disproportionately work in caring jobs, it is unlikely that policies governing wages in the care sector could achieve pay equity between men and women.

The second chapter evaluates the impact of state legislation on bullying in schools. I employ a difference-in-differences approach exploiting variation across states in the timing and type of law adopted using nationally representative surveys at the student and school levels. While I find no impact of the laws on bullying in high schools, bullying occurs most often in middle school. And impacts might vary by school type and legislation type. I also discuss current challenges to evaluating bullying legislation and provide recommendations for facilitating a conclusive assessment of whether state bullying laws work.

The third chapter uses a field experiment to evaluate an intervention aimed at increasing participation in an academic assistance program. Supplemental Instruction (SI) is a widely used, but poorly evaluated, peer-tutoring program with low participation rates. We

randomize encouragements to attend SI across a large student population. The resulting boost in participation allows us to estimate the per-session average causal impact of SI on grades for a subpopulation under certain assumptions.

EVALUATING OCCUPATIONAL OUTCOMES AND INTERVENTIONS IN SCHOOLS

BY

JULIA ANN MANZELLA

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2015

Copyright by
Julia Ann Manzella
2015

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair: Dr. Barry T. Hirsch

Committee: Dr. Paul J. Ferraro

Dr. Melissa R. Banzhaf

Dr. Rachana Bhatt

Electronic Version Approved:
Mary Beth Walker, Dean
Andrew Young School of Policy Studies
Georgia State University
August 2015

DEDICATION

To my mother, Elizabeth, and my niece, Isabella—sí se puede.

ACKNOWLEDGEMENTS

Rabah AMIR has been instrumental in my pursuit of a career as an economist. He has offered unparalleled advice without hesitation or obligation. I appreciate all of his help and thoughtfulness. Many unnamed individuals have also helped me to navigate successfully through this unimaginable journey, providing encouragement and selflessly shedding light on hazards beyond my view. I vow to pay it forward.

Thanks to all faculty, staff, and students who made my graduate experience fruitful and enjoyable. A great deal of pleasure came from playing many seasons of softball with an evolving group of talented and friendly teammates. I thank Jim Marton for providing help in subtle but important ways. Bess Blyler is truly amazing; she does everything possible to make the economics doctoral program at GSU flow more smoothly for graduate students and always with impeccable tact, professional humor, and a genuine smile. I appreciate funding received from internal awards such as the Dan E. Sweat Dissertation Fellowship, Carole Keels Scholarship in Economics, and George Malanos Economics Scholarship.

My research benefitted from helpful comments from many: my committee, Susan Averett, Paula England, David Sjoquist, Megan Skira, Aaron Sojourner, editors, referees, and various seminar/conference/workshop participants. A special thanks to Melissa Banzhaf for her careful and spot-on suggestions and insightful talks. Discussions with Tom Mroz and Spencer Banzhaf have also positively influenced my thinking and research. I would like to thank the Office of Supplemental Instruction at GSU, especially Eric Cuevas, Savannah White, and Erica Li, for the extraordinary opportunity to assist with their efforts to improve how the program is implemented; collaborating with them has been a very rewarding experience.

I am deeply indebted to Paul Ferraro for his willingness to work with me on an interdepartmental team project. To Paul, thank you for your honest feedback and advice. I have learned a lot from working with you, from critically thinking about causality to artfully employing entrepreneurial and people skills. Your gentle supervision has empowered me as a researcher.

Words cannot convey how grateful I am to have had an opportunity to work with Barry Hirsch. To Barry, thank you for enabling me to construct the dissertation of my choosing, providing generous financial support of my endeavors to become a researcher, and delivering exceptional guidance throughout the completion of my first publication—an accomplishment I am proud to share with you. I have grown so much from the validations and challenges you have extended to me. You have been an integral part of my development as a Labor Economist.

Paula Stephan is an incredible mentor, offering keen foresight, guidance, and encouragement at crucial moments. To Paula, thank you for your unconditional support.

To my parents and close friends, thank you for your patience, encouragement, and well-timed calls, cards, and packages. I am forever grateful to you, such support is priceless.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

INTRODUCTION

Since the 1950's, economists have investigated how growth in the aggregate knowledge and skills of workers may reconcile the discrepancy in economic growth that remains after accounting for growth in labor and physical capital (Becker,1975, pp. 9-10). Human capital refers to the set of knowledge, skills, and characteristics embodied in workers that contribute to their "productivity" (Acemoglu & Autor, 2015, p.3). In competitive labor markets, more productive workers earn higher wages than less productive workers. Investment in human capital helps explain economic growth.

Theory and research, however, highlight that inefficiencies arise in the labor market owing to explanations on both sides of the labor market. Labor demand factors such as wide-spread employer-based discrimination based on gender, race, or individual attributes can create substantive differences in earnings and employment opportunities for workers with otherwise comparable human capital stocks, which can in turn provide disincentives for future investments among such workers receiving lower wages. Alternatively, economists have also investigated labor supply factors for why seemingly rational individuals may not invest in themselves. Some of these explanations include imperfect information, non-cognitive skills, and behavioral responses.

Yet the evidence base is still largely incomplete as to our understanding of why some people flourish (and other do not) in the labor market, and how policy might be used to increase the likelihood of success. So with this in mind, I set out to investigate three distinct yet inter-related research questions in applied labor, each with relevance for public policy. This collection of research papers integrates a variety of topics in labor economics including wage differentials, gender and racial inequality, human capital accumulation and occupational skills. Furthermore, coupling such a scope of the field of labor with the broad set of empirical tools used herein provides a solid foundation for conducting high-quality economic analyses of policy-relevant issues. Following completion of my dissertation, I am eager to investigate research questions related to apprenticeships and policies governing them—a natural progression from evaluating occupational outcomes and interventions in schools.

Chapter I examines the role of wage penalties for caring jobs in explaining the gender wage gap. Simply put caring work involves helping and caring for others, and is done primarily by women. It is widely assumed that there exist substantial penalties for caring jobs, yet evidence and analysis supporting this is limited. All else the same, lower wages for work involving a high degree of caring for and helping others—work often done by women—can exist for reasons both consistent and inconsistent with standard theory. This paper provides a thorough analysis of wage differentials for multiple, continuous measures of caring. Detailed O*NET job descriptors matched to large, representative worker-level data sets allow us to estimate Mincerian wage equations in levels and first-differences. We find that skills matter and both women and men face caring penalties that are (perhaps surprisingly) small, on the order of about 2% for one standard deviation difference in the level of caring. While women are disproportionately employed in caring jobs, wage penalties explain little of the gender wage gap. Thus, it is unclear how recommendations aimed at wages in the care sector might provide an effective policy option for achieving pay equity between men and women.

Chapter II uses policy evaluation methods to estimate whether, and to what extent, state legislation abates bullying in schools. Such an evaluation is warranted if only to inform federal legislation that has yet to be passed even though several bills were introduced to Congress over the past decade. Other reasons make this topic interesting to economists as well as policymakers. Bullying can be financially problematic for schools. A burgeoning economics literature suggests bullying may have serious negative labor market consequences by disrupting cognitive and non-cognitive skill development. Bullying during youth may carry over into adulthood and have longer-run impacts within households, labor markets, and the larger economy. I evaluate whether bullying legislation abates bullying in schools by employing a difference-in-differences approach that exploits variation across states in the timing and type of law adopted in conjunction with nationally representative surveys of bullying outcomes at the student and school levels. Taken together, there is suggestive evidence that state legislation may have little effect on bullying in high schools, the impacts are likely heterogeneous across elementary, middle, and high schools, and the type of legislation matters. This study reveals the current challenges to evaluating bullying legislation and provides recommendations for facilitating a more conclusive assessment of whether state bullying laws work.

Chapter III involves a field experiment that uses random encouragements coupled with administrative data to estimate the causal impacts of a widely used, yet poorly evaluated, academic assistance program in higher education, namely Supplemental Instruction (SI). SI is a peer-led tutoring program that targets historically difficult courses and intends to help students master content and develop study skills. Student participation is voluntary. Proponents of SI present evidence derived from anecdotes and observational designs from which causal inferences are difficult to draw. Our research design overcomes selection problems in order to address this gap in the evidence base. Employing insights from behavioral economics, we devise an intervention that aims to encourage greater student use of SI – well-timed and targeted reminders. We then randomize this encouragement among a large student population. Using the randomized encouragement as an instrumental variable and combining our experimental data with administrative data on student characteristics and outcomes, we estimate the effects of SI on academic outcomes. In addition to shedding light on whether well-targeted reminders can boost participation in academic support programs and whether such participation improves academic performance on average, we also shed light on whether such programs can close college achievement gaps at institutions serving diverse student populations.

“Who Cares—And Does It Matter? Measuring Wage Penalties For Caring Work”
This article is © Emerald Group Publishing and permission has been granted for this version to appear here (http://scholarworks.gsu.edu/econ_diss/).
Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without express permission from Emerald Group Publishing Limited.
DOI: 10.1108/S0147-912120140000041014

I WHO CARES—AND DOES IT MATTER? MEASURING WAGE PENALTIES FOR CARING WORK

Introduction

Caring labor has been described as jobs in which workers "provide a face-to-face service that develops the human capabilities of the recipient" (England, Budig, & Folbre, 2002, p. 455).[1] Health, child, and elder care services, along with education, account for a substantial share of paid employment and personal consumption expenditures in the United States (Folbre, 2008). It is widely believed that there exist wage penalties for caring work. Research in economics and sociology provides theoretical rationales for why caring penalties can exist.[2] Yet there are surprisingly few empirical analyses examining whether such wage penalties exist and, if so, the sources of these penalties.

The approach taken by economists and other social scientists and in our study is to examine whether wages for caring jobs are high or low relative to similarly skilled workers in otherwise similar jobs and locations. The term "caring penalty" is used to mean that among workers with similar skills in similar locations working in similar jobs (apart from caring), lower wages are found in jobs requiring higher levels of caring. In previous work, England et al. (2002) find overall wage penalties for caring labor in the United States. They find that the type of care work matters, with nurses enjoying a significant wage premium and workers in most other caring occupations suffering penalties. Their study, as well as others in the literature, assumes a

---

[1] England et al. refer to "human capabilities" as "health, skills, or proclivities that are useful to oneself or others." "Caring labor" refers to jobs (occupations) that require caring tasks. The literature attempts to estimate wage differences among jobs that involve high and low levels of caring tasks. This is not necessarily the same thing as wage differences between workers who do and do not have caring attitudes and behaviors. Labor market sorting no doubt results in caring persons working disproportionately in jobs with caring tasks.

[2] Works include England and Folbre (1999), Folbre and Nelson (2000), England (2005), and Folbre (2006, 2008, 2012). There exists a separate literature focusing on unpaid informal care and how it affects caregivers' labor force participation and wages. See Van Houtven, Coe, and Skira (2013).

dichotomy in care work with occupations classified as either involving or not involving a high degree of care.

The question of whether there exist significant wage penalties for caring work is important for several reasons. Depending on the source, wage penalties for care work might be viewed as an equity problem if the incidence of such penalties disproportionately affects women (or other groups). And penalties might be viewed as a social and economic problem if low wages in the care sector create higher than optimal turnover and low quality care in socially valuable jobs (England et al., 2002). Moreover, a finding of sizable wage gaps among truly similar workers in similar jobs (apart from the degree of caring) can raise the question of whether labor market outcomes deviate substantially from what is predicted by standard theory, depending on the source of these differentials.

Our paper provides evidence intended to enhance knowledge about wage differences associated with caring. In what follows, we first discuss how standard theory might account for wage penalties for caring work. We then provide detailed empirical analysis on how wages differ across workers and jobs with respect to caring attributes. To do so, we match employee data from Current Population Survey (CPS) earnings files with detailed occupational job descriptors, including multiple measures of caring, from the 2007 Occupational Information Network (O*NET). Cross section analysis is conducted using large CPS data files for 2006-2008. Longitudinal analysis is conducted using large CPS panels for worker-year-pairs from 2003/2004 to 2008/2009, with each pair consisting of two observations per worker, one year apart. The panel analysis identifies differentials for caring work based on wage changes among job switchers who increase or decrease the required levels of caring in their jobs. As compared to prior literature, our analysis provides more recent evidence, uses large cross-sectional and panel samples of workers, provides multiple continuous (rather than categorical) measures of caring (and other) job attributes, and examines wage differentials associated with these measures in a comprehensive fashion.

Theoretical Rationales

There is a literature in sociology and economics in which researchers propose theories that might explain wage penalties for care work.[3] The mechanisms emphasized by sociologists, while not necessarily described using an economic framework, are largely compatible with economic theory once framed in language familiar to economists. In the remainder of this section, we use standard theory to discuss how systematic wage differentials for caring work may arise in the labor market.

The most obvious explanation for wage differentials associated with caring work is the theory of compensating differentials, whose sterling pedigree (i.e., Adam Smith's Wealth of Nations) is unrivaled. Some individuals will derive greater utility from work characterized by a high degree of caring for and helping others than from work that is not. And these workers may be disproportionately female. If such preferences are sufficiently widespread so that they are relevant at the margin (i.e., where labor supply and demand intersect) and not inframarginal, theory suggests that jobs involving high levels of caring will bear a wage penalty compared to jobs with otherwise similar working conditions and skills.

Compensating differentials are illustrated in Figure 1, which shows labor demand and supply for an occupation that involves a high level of caring. To illustrate our point, we show a

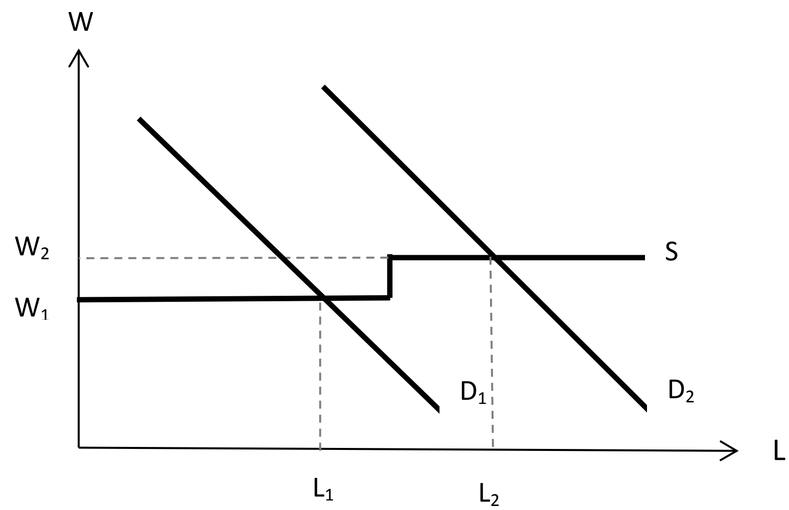
[3] See England and Folbre (1999), England et al. (2002), England (2005), and Barron and West (2013).

diagram in which workers are equivalent except that one group prefers work in a caring job, while other workers are indifferent to (or dislike) caring tasks. The lower portion of the supply curve SL represents occupation labor supply for the first group of workers, whose reservation wage is W1. The upper portion of SL represents labor supply for the second group, whose reservation wage is W2. In this simple case, if labor demand in the occupation is relatively low at D1, we will see an equilibrium with wage W1 and employment L1, the latter including only workers who prefer caring jobs. At a high level of demand D2, the equilibrium wage is W2, a higher wage for both groups of workers. The more general case is that there is a continuum of preferences among workers leading to an upward sloping SL.[4] The lower the level of demand, the lower the wage and thus "penalties" (negative compensating differentials) exist for caring work. At high levels of demand, "penalties" should decrease and eventually disappear with sufficiently high demand. If demand is high and workers at the margin dislike caring work, then we should see a wage "premium" (positive compensating differentials).

Researchers have referred to this "preference" argument as the "intrinsic rewards" or "prisoner of love" explanation (England, 2005; England & Folbre, 1999, respectively). Economists are inclined to treat preferences as more or less freely chosen and deserving to be respected in markets. The concern is that selection into caring jobs may not fully reflect free choice, being seen instead as an obligation driven by societal expectations (hence the "prisoner of love" label).

[4] In labor textbooks (e.g., Borjas, 2013, p. 213), the relationship between wages and job attributes is typically shown as a hedonic wage function formed by the tangencies of workers' indifference curves and firms' isoprofit curves, an approach developed by Thaler and Rosen (1976).

Figure 1. Labor Market Equilibria for Caring Occupations with Heterogeneous Preferences



W_1 represents the reservation wage for workers who prefer caring jobs and W_2 the reservation wage for those without such a preference. At low labor demand level D_1 the equilibrium wage for the caring job is W_1 with employment L_1. At the high level D_2 the equilibrium wage for both sets of workers is W_2 with employment L_2.

"Devaluation" is a prevalent theory for a caring wage penalty emerging from sociology. It makes the assumption that society values whatever is produced by women less than what is produced by men; hence, caring labor pays less because women disproportionately work in caring jobs (England et al., 2002). A devaluation explanation can be reconciled with standard micro theory to the extent that devaluation means that individuals, in their roles as consumers, employers, voters, etc. have a lower willingness to pay for services typically provided by women. Combining lower labor demand with upward sloping long-run labor supply curves due to heterogeneous skill and job preferences can then produce wage differences between women and men and caring versus noncaring work.[5]

[5] Although gender discrimination in the workplace is illegal, lower wages resulting from labor supply preferences and "devalued" work would not typically be illegal.

Independent of women's preferences for caring work, if there exists substantial employer-based discrimination against women in higher-paying noncaring jobs, women may be "crowded" into caring sectors, lowering equilibrium wages in caring jobs and denying women opportunities to accumulate human capital outside that sector. This characterization of the U.S. labor market may nicely fit the first three-quarters of the twentieth century, but over the last 40 years or so it has become less accurate.[6]

An additional argument in the literature is that caring labor is disproportionately concentrated in the public sector (Barron & West, 2013; Folbre, 2012), an outcome that we will subsequently document. If the political process produces a relatively low willingness to pay for public services that most involve caring, then low wages can result. In our empirical work, we separately examine how wage outcomes vary with respect to caring in the private and public sectors.

Prior studies have identified reasons one might see lower wages in caring jobs. One can also identify channels that might produce higher wages in caring jobs, or at least mitigate penalties. As mentioned in our discussion of compensating differentials, if demand for caring jobs is sufficiently high such that the marginal worker receives disutility for caring tasks, then we should see positive rather than negative compensating differentials. Efficiency-wage theory and theories of reciprocity and gift exchange suggest that caring jobs might have high rather than low wages. Measuring and monitoring the quality of care services provided by an employee can be difficult for employers or customers (e.g., parents selecting child care), so high wages may arise to attract more-able workers, reduce shirking, lower turnover rates, and foster

[6] Women were also crowded into many low-pay, low-caring clerical and factory jobs.

gift-exchange between employees and the employer/ customer (Borjas, 2013, pp. 484-493; Fehr & Gachter, 2000). It seems plausible to us that such forces might well mitigate to some (unknown) degree negative wage effects resulting from worker labor supply (i.e., intrinsic rewards for care work) and demand-side devaluation of caring jobs.

A final consideration is whether we should expect different wage-caring gradients for women and men, given that each group's selection into caring jobs can differ. Women bear a disproportionate share of caring, assisting, and teaching others within households. The acquired skills, preferences, and societal norms that accompany such a household division of labor make it more likely that in the labor market women sort into jobs requiring similar or complementary skills to those provided in the home. If women (on average) are particularly adept at caring tasks, we may observe smaller wage penalties for caring work among women than among men (i.e., biased estimates due to unmeasured productivity). A related argument stems from recent work by Heckman (e.g., Heckman & Kautz, 2012) regarding the importance of noncognitive skills (e.g., conscientiousness) on success in the labor market and elsewhere. If women (on average) have higher levels of "people" skills (see Borghans, ter Weel, & Weinberg, 2008, 2014) valued in jobs involving high levels of caring, these unmeasured person-specific skills should lead to higher wages and weaker estimates of caring wage penalties for women than for men using cross-sectional analysis.[7] As discussed subsequently, panel estimates, herein

[7] Borghans et al. (2008) refer to caring as an interpersonal style that is more aligned with the notion of interactive service work discussed in England and Folbre (1999).

caring wage differentials are measured by wage changes among workers switching between high and low caring jobs, should account for these person fixed effects.[8]

Previous Evidence

There is a limited number of empirical studies providing in depth analysis of caring wage differentials. We summarize two studies that are most similar to our work in terms of the research question being addressed. England et al. (2002) analyze the relative pay of caring labor in the United States using individual longitudinal data from the NLSY79 for the years 1982-1993. Their sample consists of 10,670 respondents for whom they had at least two years of detailed employment data. Respondents were ages 16-23 in the initial year of their sample (1982) and 28-35 in the final year (1993).[9] The authors create an indicator variable designating caring occupations, these being primarily in health care or education, plus a handful of other occupations (childcare workers, librarians, counselors, social workers, clergy and other religious workers, and recreation and fitness workers). An advantage of their data set is that they are able to construct measures of previous part-time and full-time work experience, job tenure, and breaks in employment. Such measures are particularly important for cross-sectional analysis, but effectively fall out in longitudinal analysis (through worker fixed effects or wage change analysis), the method favored in their study and in our subsequent analysis. In order to control for occupational skill, they use factor analysis to create a measure of cognitive skills

[8] We do not explore the possible effect of monopsonistic labor markets, where a limited number of employers and little worker mobility across firms leads to lower wages. As discussed in Manning (2003) and Webber (2013), if women have relatively less job-to-job mobility than do men, their wages are likely to be lower. Although empirical tests of monopsony models might help explain lower wages for women than men they do not necessarily explain lower wages for caring than non-caring jobs, absent evidence that job-to-job mobility is lower in high caring than in low caring jobs.

[9] They restricted their sample to respondents who worked either part-or full-time for at least two years during the sample period (person-years with missing values or extremely low or high hourly earnings were also dropped).

demanded by an occupation based on job descriptors in the Dictionary of Occupational Titles.[10] Their fixed-effects wage regression models, with dummies for individuals and years, identify the wage effect of caring based on individuals who switch between care and noncare occupations. They perform their analysis separately for men and women, finding significant wage penalties of 5% for women and 6% for men.

When England et al. (2002) disaggregate care work into seven types, they find large heterogeneity in the estimated wage gaps. Female childcare workers were found to face large penalties. Surprisingly, doctors were found to suffer substantive wage penalties (men: 17%; women: 10%) while the category containing nurses, therapists, and medical assistants, referred to as “other medical,” enjoyed a wage premium (men: 4%; women: 8%). The result for doctors is counterintuitive, but likely arises from age-truncation in the sample, as noted by the authors (England et al., 2002, p. 468). The result regarding nurses is consistent with subsequent nursing studies, although Hirsch and Schumacher (2012) show that such estimates overstate relative nursing wages.[11]

The data and analysis in England et al. (2002) have advantages and limitations as compared to our work. The principal advantage is the NLSY longitudinal structure and the detailed information available on individuals and households over time. That said, the NLSY79 sample period covered 1983-1992, so all workers were under age 36 and were observed early

[10] The DOT was the precursor to O*NET, the latter being used in our analysis to construct job attribute indices using factor analysis. While the DOT included a small number of job descriptors with a limited range of integer coding, O*NET contains several hundred job skill/ task and working condition attributes, each providing non-categorical continuous ratings. O*NET is discussed subsequently in our data section.

[11] Hirsch and Schumacher (2012) show that nursing/non-nursing wage gaps shrink substantially once one: (a) controls for occupation skill requirements and working conditions and (b) accounts for bias in estimating log (percentage) wage differentials with OLS when comparing treatment groups with low wage dispersion (nurses) to broad comparison groups with much higher dispersion (all college-educated women). See Blackburn (2007).

in their careers, as noted by the authors. Their period of analysis, roughly twenty years earlier than in our study, may not reflect the experience of more recent cohorts, with occupational shifts among women due to evolving social norms and increased educational attainment, or the substantial changes in overall labor market rewards to skill and the skill content of jobs. And our CPS panels provide relatively larger samples of occupation switchers than does the NLSY, having roughly 19 and 22 thousand unique worker job switchers for women and men, respectively.

In a recent study, Barron and West (2013) analyze wage differentials for caring work in the British labor market. They use 17 waves of the British Household Panel Survey covering the years 1991-2007, consisting of annual accounts of an individual’s education, employment, and family characteristics. They estimate a regression model that includes individual random effects (not fixed effects) and year fixed effects. Similar to England et al. (2002), Barron and West disaggregate care work into six specific types (doctors, nurses, teachers, childcare, nursing assistant, and welfare) by using self-reported job descriptions matched to occupation codes. Their empirical work combines men and women (a male dummy is included), thus not allowing for different caring penalties (rewards) for women and men. Barron and West rely on a Heckman selection correction procedure to account for the attrition of low-wage workers from the labor force, but do not discuss what exclusion restriction (if any) was used or how their selection results compare with OLS, thus making interpretation difficult. The authors construct different comparison groups based on broad occupation (described as socioeconomic groups) for each group of care workers.

Barron and West obtain wage differential estimates indicating that doctors earn 33% more, nurses 10% more, and teachers 5% more than their peers in noncaring occupations, while nursing assistants earn 6% less, welfare workers 18%less, and childcare workers 21% less than their peers in noncaring occupations. Unlike England et al. (2002) and our subsequent analysis, Barron and West do not include measures of occupational skills, nor do they account for worker fixed effects. We suspect that their large caring wage gap estimates (negative and positive) may reflect substantial skill differences between their caring and noncaring occupations and workers. And, as the authors point out, important differences exist between the U.S. and U.K. labor markets, in particular, the U.K.'s nationalized health care and education systems in which wages are set by national negotiations (Barron & West, 2013).

Data Sources and Descriptive Evidence on Caring Jobs

The data sets used in our analysis are constructed from two principal sources, the Current Population Survey Outgoing Rotation Group (CPS-ORG) monthly earnings files, containing worker-level data, and the Occupational Information Network (O*NET), providing occupational descriptors that can be matched to the CPS.[12] A cross-sectional data set is created by pooling 36 monthly CPS-ORG files for January 2006 through December 2008, with 166,009 women and 168,760 men included in the estimation sample. Since the CPS includes households in the same month in two consecutive years, we are able to construct a large panel data set

[12] The CPS-ORG data is jointly sponsored by the U.S. Census Bureau and the Bureau of Labor Statistics (BLS) while O*NET is sponsored by the Department of Labor's Employment and Training Administration (USDOL/ETA).

consisting of about forty-thousand worker-year pairs of occupation switchers from 2003/2004 to 2008/2009 of which 18,981 are women and 21,689 men.[13]

An occupational-level data set constructed from a 2007 edition of O*NET provides several hundred job attributes. Most O*NET variables provide ratings of the skills and various tasks required to perform the job or the environment of a worker in the job. We use 206 O*NET variables, most measured on scale indicating the level to which a descriptor is required or needed to perform the occupation. Most O*NET variables are measured on a scale from 0-to-7 or 1-to-5, with reported values being a continuous number based on ratings provided by job analysts based on site visits and reports from job incumbents.[14] We scale all O*NET measures from zero to one to provide a comparable scale.[15]

In line with prior literature, we regard care work as involving activities and requirements in jobs characterized by high levels of nonroutine interactive job tasks that directly foster recipients' social, emotional, intellectual, and/or physical well-being. Typically, the delivery and quality of caring job skills/tasks depend on workers providing individualized services and

[13] The time period for the panel was determined by there being time-consistent detailed Census occupation codes in the CPS for the years 2003-2009. For details on the methods used to match individuals across years in CPS earnings files, see the appendix in Macpherson and Hirsch (1995) and Madrian and Lefgren (2000). Matching is conducted with the goal of including only pairs matched with near certainty (using household and person identifiers and demographic checks), even if it means excluding valid pairs that do not satisfy all match criteria. In addition, we exclude all individual observations with imputed earnings in the pooled cross section analysis and all earnings pairs with either or both years imputed in the panel analysis. In the CPS-ORG files, earnings non-respondents are assigned the earnings of a "similar" donor based on broad but not detailed occupation. Hence, the estimates of coefficients on caring or other occupation-based variables will be attenuated (so-called "match bias") if imputed earners are included. See Bollinger and Hirsch (2006) for a detailed discussion of CPS imputation methods, resulting biases, and alternative corrections for such biases. Simply omitting imputed earners avoids imputation match bias and provides estimates highly similar to more complex correction methods.
[14] We use the O*NET data set previously used in Hirsch and Schumacher (2012). They provide a more detailed description of its construction and merger with the CPS, Section4of their paper provides discussion of the selection of variables used to construct the skill and working condition indices. Most O*NET descriptors are measured by both required level and importance. The two rankings are highly collinear; we use levels.
[15] In order to normalize ratings for attributes using a different scale, we follow an approach similar to the one used by the USDOL; namely score $S = \frac{O-L}{H-L}$, where O is the original rating score on the rating scale used, H is the highest possible score on the rating scale used, and L is the lowest possible score on the rating scale used.

establishing a personal relationship with the recipient.[16] Consequently, we identify two O*NET occupational attribute variables as most directly measuring the level of caring work: "assisting & caring for others" (A&C) and "concern for others." We examine these O*NET caring measures separately and jointly, using factor analysis to construct a latent factor combining the two measures, which we refer to as the "caring" index. Table 1 provides the definition for the O*NET caring and developing/ teaching attributes described below.

The "assisting & caring for others" variable is designed to measure job requirements. A&C is included in a category of O*NET attributes labeled "communicating and interacting with others," which is one of the multiple categories under the broader category of "generalized work activities," all part of the larger content category "occupational requirements." The "concern for others" measure, on the other hand, falls under the broad content category "worker characteristics," with a sub-heading of "work styles" and the more narrow sub-heading "interpersonal orientation." The "A&C" and "concern" measures have a high degree of statistical overlap. That said A&C is a job descriptor designed to measure the level of required caring work activity in an occupation, whereas "concern" is designed as a worker descriptor that identifies personal characteristics needed for job performance and for a good occupational match.

In addition to the caring and concern measures, we construct a "developing/teaching" or D/T factor index, which loads four O*NET descriptors: "developing & building teams,"

[16] Our definition is similar to England et al. (2002), but also incorporates the concepts of "routine versus nonroutine" and "interactive" job tasks developed in Autor, Levy, and Murnane (2003) in their analysis of information technology (IT) on employment and wages. Each of these studies uses DOT occupation job descriptors to classify jobs.

Table 1. O*NET Job Level Attributes Describing Care Work

	O*NET attribute label	O*NET attribute description	O*NET variable	Possible score range
Caring	Assisting and caring for others	Providing personal assistance, medical attention, emotional support, or other personal care to others such as coworkers, customers, or patients	4.a.4.a.5	[0, 7]
	Concern for others	Job requires being sensitive to others' needs and feelings and being understanding and helpful on the job	1.c.3.b	[1, 5]
Develop/Teach	Developing and building teams	Encouraging and building mutual trust, respect, and cooperation among team members	4.a.4.b.2	[0, 7]
	Training and teaching others	Identifying the educational needs of others, developing formal educational or training programs or classes, and teaching or instructing others	4.a.4.b.3	[0, 7]
	Coaching and developing others	Identifying the developmental needs of others and coaching, mentoring, or otherwise helping others to improve their knowledge or skills	4.a.4.b.5	[0, 7]
	Instructing	Teaching others how to do something	2.b.1.e	[0, 7]

Source: Author created using information from http://www.onetonline.org/. Original score range refers to the minimum and maximum score values in our sample before rescaling all O*NET attributes on [0, 1] using the formula $S = \left(\frac{O-L}{H-L}\right)$, where *O* is the original rating score on the rating scale used, *H* is the highest possible score on the rating scale used, and *L* is the lowest possible score on the rating scale used. The caring variables are normalized to mean zero and s.d. equal 1 when included in wage regressions (column 4 of the regression tables) so coefficients can measure the log wage effect of a one s.d. change in the O*NET measure.

“training & teaching others,” “coaching & developing others,” and “instructing.” As was the case for the A&C measure, the four D/T descriptors describe required occupational work activities. The D/T measures are not intended to directly measure caring wage effects. Because prior studies typically designate teaching jobs as caring jobs, it is important that we separately examine the relationship between wages and D/T tasks as well as wages and caring tasks. The advantage of our approach is that all occupations are included, each with distinct measures of levels of caring and D/T, allowing us to estimate separate wage gradients with respect to each.

In addition to the job caring indices, we construct two broad factor indices, one reflecting occupation job skill and task requirements and a second reflecting job working conditions.[17] Our “job skills index” includes 162 O*NET job skill/task variables and heavily loads cognitive skills. For example, heavily-loaded attributes include levels of critical thinking, judgment and decision making, monitoring, written expression, speaking, active listening, active learning, negotiation, and persuasion. A second factor index -a “working conditions index” is constructed by loading 38 O*NET variables measuring (mostly) physical working conditions. This index heavily loads attributes such as required types of strength, extreme temperatures, extremely bright or inadequate lighting, exposure to contaminants, cramped work space or awkward position, exposure to injuries, and exposure to hazardous equipment.

The skills index is a particularly important control variable in the wage analysis for two reasons. First, the job skills index is a strong correlate of wages and, second, caring jobs vary greatly in their required level of skill, some involving minimal training and low levels of cognitive skills, while others are among the most highly skilled jobs in the economy. Caring jobs

[17] Our approach in forming the O*NET skills and working condition indices follows Hirsch and Schumacher (2012). For further details on the O*NET to CPS match, see their paper.

also vary with respect to working conditions, although these attributes have far weaker effects on wages than do skills. As recognized in the literature, it is difficult to identify compensable working conditions due to (a) heterogeneous preferences and sorting with respect to job attributes and (b) because job disamenities tend to be negatively correlated with unmeasured worker skills (Hwang, Reed,& Hubbard, 1992).

The O*NET attributes and factor indices measuring occupation skills, working conditions, and the various measures of caring are matched by detailed occupation to both the 2006-2008 CPS cross-sections and the 2003/2004-2008/2009 CPS panel data sets. Table 2 provides the unweighted means and standard deviations for women and men for the O*NET measures and indices, plus the earnings measure used in the CPS, as described below.[18]

Our earnings measure is the natural log of average hourly earnings across all hours worked, with earnings inclusive of tips, overtime, and commissions, for individual worker i, in constant 2008 dollars. In our regression models (shown subsequently), we include all nonstudent workers ages 18-65 with hourly wage values between three and one hundred fifty dollars. We exclude full-time students (reported for those under age 25) and observations in which workers' earnings are not reported and instead imputed by Census, since workers' detailed occupation is not a hot deck match attribute used to assign donor earnings to nonrespondents (Bollinger & Hirsch, 2006, Table 1). The same sample exclusions are applied to the panel analysis covering 2003/2004-2008/2009, where the dependent variable is the one- year change in the log of average hourly earnings. Were imputed earners included, coefficient estimates regarding caring attributes in the wage level analysis would be attenuated, whereas panel

[18] Differences between sample-weighted and unweighted descriptive statistics and regression analyses are trivial.

estimates would be severely biased toward the cross-section results (see Bollinger& Hirsch, 2006).[19]

Turning to Table 2, the raw gender gap is 0.19 log points (roughly 20%), with women's mean hourly earnings $4.38 less than that for men and somewhat less dispersed.[20]20 The other variables in Table 2 are occupation measures, with the means compiled across workers (i.e., equivalent to a sample-weighted mean across occupations). We use six "caring" attributes from O*NET, measured in levels and each scaled between 0 and 1. The two attributes used to measure caring are "assisting & caring for others" and "concern for others." Women have higher averages than do men for each, 0.46 versus 0.39 for the former and 0.78 versus 0.69 for the latter. Our "caring" factor index that combines these two attributes has a substantially higher value for women than men, the difference exceeding a half standard deviation.

Table 2 also presents means for the four occupational measures emphasizing aspects of team development, training, coaching, and instructing. Here, women and men have highly similar levels for each. Even if these job attributes were associated with substantial differences in wages, this would produce minimal changes in the gender wage gap. We combine these four O*NET attributes into the factor index labeled "developing/ teaching" (D/T).

---

[19] Inclusion of imputed earners does not correct for (possible) non-ignorable response bias, since included non-respondents are assigned the earnings of respondents. One can reweight the respondent sample by the inverse probability of response, which rebalances the sample based on measured attributes. But in practice IPW regression results are nearly identical to those using unweighted respondent samples. See Bollinger and Hirsch (2006, 2013).

[20] Throughout the paper we will treat log wage gaps as approximate percentage differentials, with the implicit wage base being in between the average for women and men (roughly the geometric mean).

Table 2. Summary Statistics for CPS and O*NET Attributes, by Gender, 2006-2008

	Women		Men		
Variables	mean	s.d.	mean	s.d.	$\bar{X}_F - \bar{X}_M$
Hourly earnings	18.83	13.51	23.21	17.59	-4.38
Ln(hourly earnings)	2.7635	0.5602	2.9531	0.5931	-0.1896
Usual hours worked per week	37.2	9.7	42.2	9.3	-5.0
Individual O*NET attributes (scaled 0-1):					
Concern for others	0.7793	0.1124	0.6916	0.1102	0.0877
Assisting & caring for others	0.4576	0.1625	0.3869	0.1281	0.0707
Developing & building teams	0.3681	0.1321	0.3854	0.1281	-0.0173
Training & teaching others	0.4269	0.1358	0.4249	0.1219	0.0020
Coaching & developing others	0.4331	0.1510	0.4274	0.1450	0.0057
Instructing	0.5822	0.1252	0.5633	0.1088	0.0189
O*NET Indices (using factor analysis):					
Caring index (loads concern and asst/caring above)	0.2950	0.8748	-0.2622	0.7420	0.5572
Developing/Teaching index (loads other 4 attributes)	0.0722	1.0062	0.0540	0.9201	0.0182
Job skills index (loads 162 O*NET attributes)	0.1157	0.9632	0.0332	1.0161	0.0825
Working conditions index (loads 38 O*NET attributes)	-0.4124	0.6790	0.3554	1.1155	-0.7678
Sex composition (%Female, from CPS)	0.6761	0.2332	0.3074	0.2468	0.3687
N	166,009		168,760		

Variable means and s.d. created from the CPS are unweighted (weighted means are highly similar). Hourly earnings are measured using implicit hourly wage (usual weekly earnings/usual weekly hours worked). All indices formed from factor analysis are compiled using the combined female and male sample and, by construction, have mean 0 and s.d. 1.0. The job skills index loads 162 O*NET job attributes and does not include the 6 O*NET 'caring' attributes. See the text for further details.

To get a feel for how various occupations are rated by O*NET with respect to "assisting & caring for others" and "concern for others," in Table 3 we provide ratings for selected occupations that have very high and low ratings, plus many of the larger occupations in the economy. What can be clearly seen in Table 3 is that many of the highest ranked occupations are health care jobs in which workers directly interact with individuals. Notable among the occupations ranked one-to-five in "assisting & caring" (physician assistants, physicians and surgeons, LPN/LVNs, respiratory therapists, and registered nurses), are the substantial differences in required skills and pay. This reinforces our previous statement that in order to measure wage differences associated with caring, it is essential to have good controls for individual worker skills and job skill requirements. Occupations requiring minimal levels of "assisting & caring" include engineers, mathematicians, machinists, and sales representatives. Examining the rankings for "concern for others" shows that there is a strong correlation with "assisting & caring" but that an occupation can be ranked high (or low) using one measure but not the other. The sales representatives occupation involves few "assisting & caring" tasks, but "concern for others" is a worker attribute that is helpful in performing a sales job. Ambulance drivers are ranked 8th highest in "assisting & caring" but 71st in "concern" for others.

A principal takeaway from Table 3 is that all jobs require tasks involving some degree of "assisting & caring" and that adequate performance in all jobs requires that workers have some degree of "concern" for others. Characterizing occupations as either caring or not provides a useful shorthand for discussion. But in order to statistically estimate the relationship between market wages and caring, it makes sense to examine multiple caring measures and explicitly account for the required levels of care for each. Rather than estimating wage differentials

Table 3. Measures of Caring for Occupations with High or Low Caring Ratings and Selected Large Occupations

COC	Occupation Name (Standard Occupational Classification)	A&C rank	A&C value	Concern rank	Concern value
3110	Physician assistants	1	0.961	2	0.983
3060	Physicians and surgeons	2	0.961	12	0.945
3500	Licensed practical nurses & licensed vocational nurses	3	0.886	6	0.970
3220	Respiratory therapists	4	0.886	22	0.923
3130	Registered nurses	5	0.876	13	0.943
9110	Ambulance drivers & attendants, exc. emergency med. tech.	8	0.823	71	0.835
2040	Clergy	12	0.799	103	0.800
3160	Physical therapists	15	0.781	4	0.978
3850	Police and sheriff's patrol officers	18	0.771	75	0.831
3640	Dental assistants	22	0.743	73	0.833
3630	Massage therapists	36	0.679	7	0.965
2010	Social workers	40	0.666	25	0.921
3650	Medical assistants and other healthcare support occupations	45	0.644	60	0.859
4600	Child care workers	46	0.642	34	0.898
2310	Elementary and middle school teachers	73	0.571	19	0.931
3600	Nursing, psychiatric, and home health aides	79	0.554	54	0.865
620	Human resources, training, and labor relations specialists	104	0.525	59	0.860
2200	Postsecondary teachers	115	0.500	121	0.783
5700	Secretaries and administrative assistants	153	0.433	116	0.790
4110	Waiters and waitresses	165	0.420	101	0.745
5240	Customer service representatives	190	0.403	117	0.785
4720	Cashiers	195	0.401	152	0.754
1340	Biomedical engineers	202	0.393	272	0.670
120	Financial managers	253	0.366	301	0.651
9130	Driver/sales workers and truck drivers	266	0.359	341	0.623
9620	Laborers and freight, stock, and material movers, hand	270	0.356	402	0.583
10	Chief executives	293	0.346	176	0.738
5620	Stock clerks and order fillers	299	0.343	265	0.673
4760	Retail salespersons	301	0.341	183	0.733
2100	Lawyers, Judges, magistrates, and other judicial workers	310	0.337	349	0.620
7200	Automotive service technicians and mechanics	332	0.328	419	0.574
6230	Carpenters	355	0.315	385	0.600
5120	Bookkeeping, accounting, and auditing clerks	367	0.306	227	0.700

(continued on next page)

Table 3. (continued)

4220	Janitors and building cleaners	377	0.301	167	0.745
4020	Cooks	390	0.294	260	0.676
800	Accountants and auditors	400	0.287	329	0.630
1020	Computer software engineers	457	0.228	413	0.577
8030	Machinists	471	0.211	500	0.380
4850	Sales representatives, wholesale and manufacturing	472	0.210	203	0.720
1300	Architects, except naval	490	0.176	473	0.506
1210	Mathematicians	495	0.143	501	0.333
300	Engineering managers	499	0.103	308	0.645
1320	Aerospace engineers	501	0.019	373	0.613

'A&C' and 'Concern' denote O*NET attributes 'assisting & caring for others' and 'concern for others', respectively. COC denotes the 2002 Census occupation codes adopted in the CPS beginning in 2003. Value refers to the rescaled score of an occupation's corresponding O*NET attribute ranking on a [0,1] scale. See details in the note to Table 1.

between jobs designated as caring or not, we estimate labor market wage gradients with respect to the degree of caring across all occupations.

Table 4 reports the simple pairwise correlations between earnings, gender, and the various O*NET job descriptors and indices. The O*NET descriptors "assisting & caring for others" and "concern for others" have a 0.71 correlation. The caring index combining these two caring measures is positively correlated with our comprehensive job skill index (0.39), negatively correlated with the index of physical working conditions (−0.21), positively but weakly correlated with the log wage (0.07), and strongly correlated with the share of women in an occupation (0.54). The developing/teaching index is positively correlated with wages and the caring measures, but largely uncorrelated with gender.

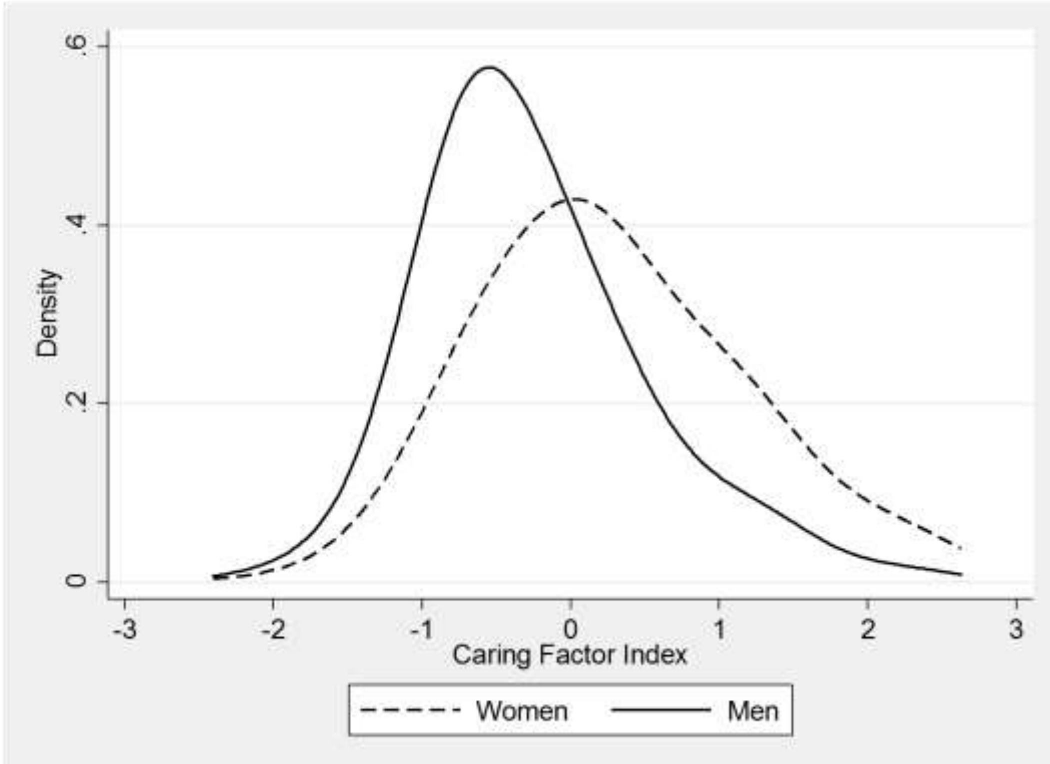
Figure 2 shows smoothed distributions of the indices for caring and developing/teaching across the labor market (absent smoothing one observes multiple "mini-peaks" at index values attached to large occupations). Recall that by construction, the means of the factor indices,

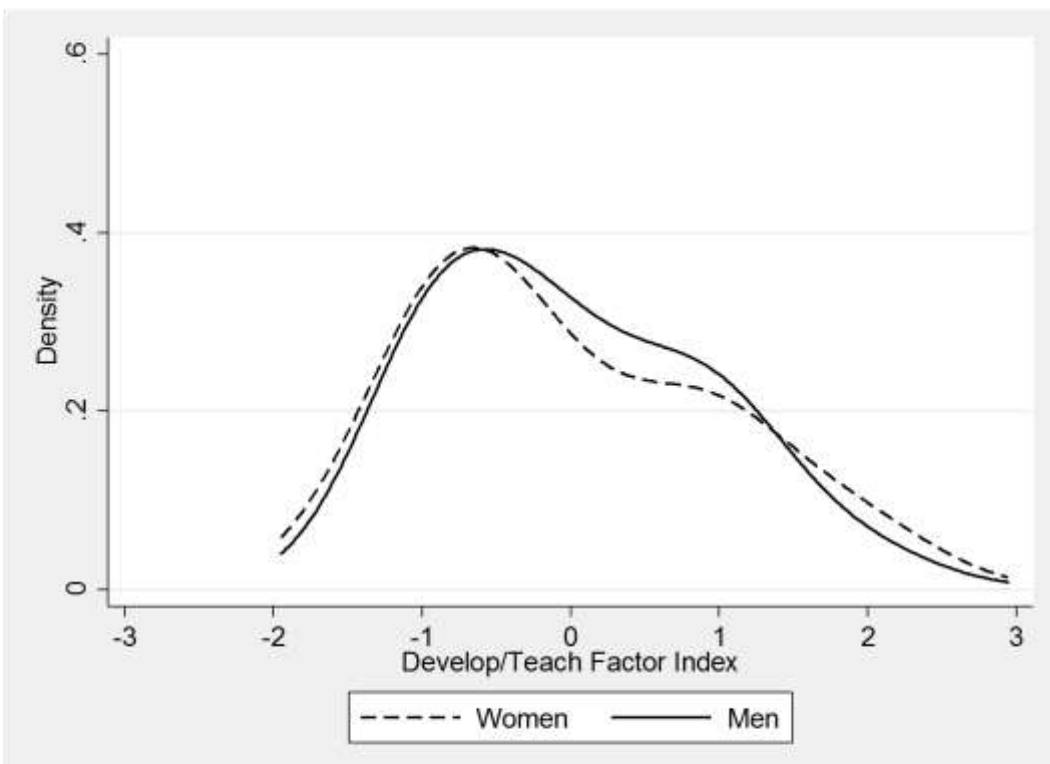
Table 4. Pairwise Correlations of Earnings, Gender, and O*NET Occupational Attributes

	Ln(hourly earnings)	Female	Occupation %Female	Concern For others	Assist and Caring	Caring factor index	Develop/ Teach index	Job skills index	Work cond index
Ln(hourly earnings)	1.0								
Female	-0.16	1.0							
Occupation %Female	-0.13	0.61	1.0						
Concern for others	0.04	0.37	0.60	1.0					
Assisting & caring for others	0.08	0.24	0.39	0.71	1.0				
Caring factor index	0.07	0.33	0.54	0.93	0.93	1.0			
Developing/Teaching (D/T) factor index	0.35	0.01	0.02	0.37	0.49	0.47	1.0		
Job skills factor index	0.53	0.04	0.08	0.36	0.36	0.39	0.75	1.0	
Working conditions factor index	-0.20	-0.38	-0.63	-0.32	-0.08	-0.21	-0.20	-0.49	1.0

These correlations are calculated using a pooled sample of 36 CPS-ORG monthly earnings files for female and male workers covering January 2006 to December 2008 merged with O*NET job attributes. Sex composition is calculated as the ratio of the mean number of females to mean total workers in a given occupation. Observations are not weighted by CPS sampling weights. Hourly wage is an implicit measure based on worker self-reports of usual weekly earnings divided by usual weekly hours worked. See text for details.

Figure 2. Density Plots of O*NET 'Caring' and 'Developing/Teaching' Factor Indices



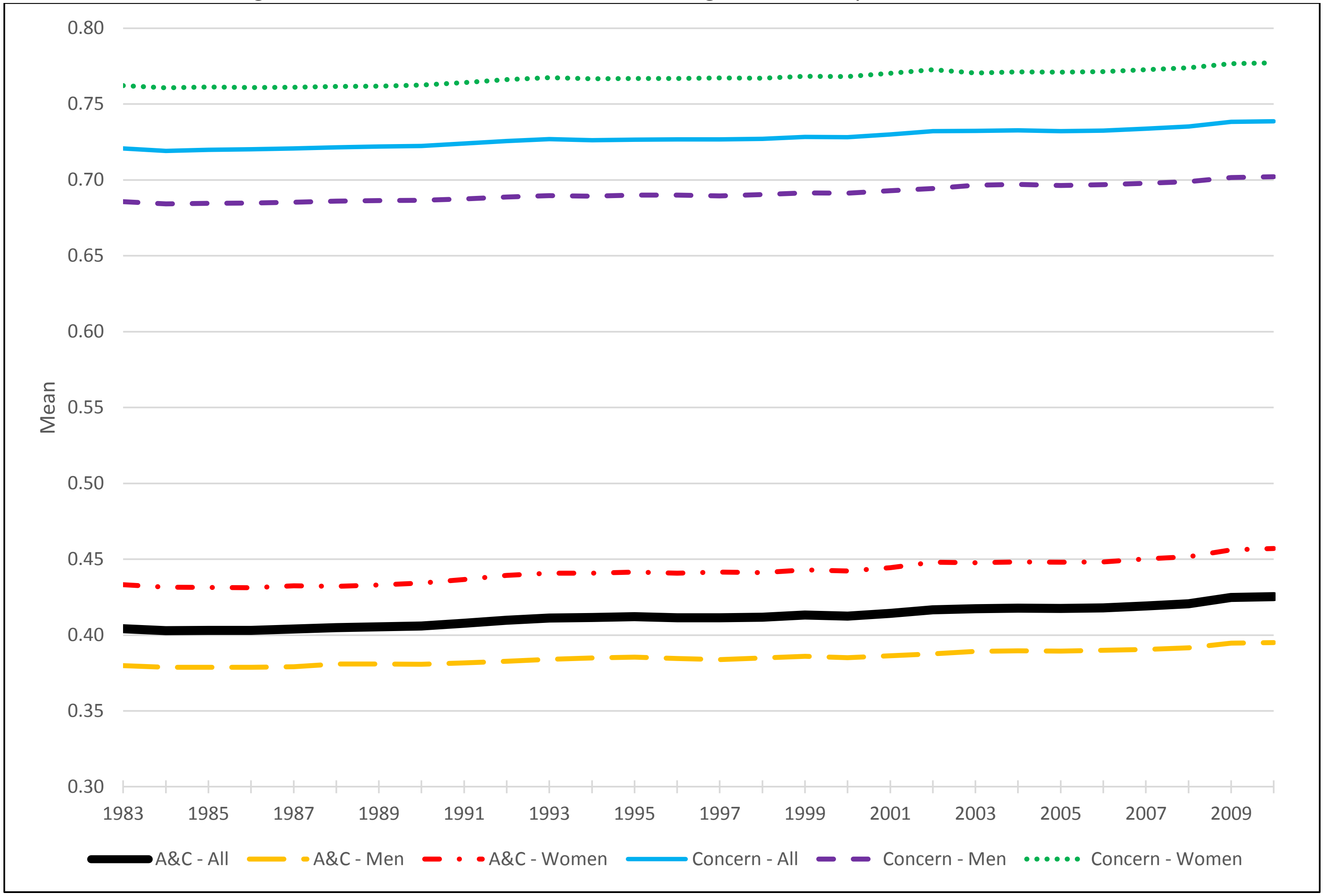


constructed for women and men combined, have a zero mean and s.d. of 1.0. In Figure 2, the caring index distribution for women is everywhere to the right of the male distribution, demonstrating both higher and more dispersed levels of job caring than seen among men. By contrast, the D/T index, which combines the four pertinent O*NET measures, displays differences between women and men, but is not systematically higher for one group or the other. The left tails and peaks of the two distributions are similar, but beyond the peaks, men are more heavily represented in occupations requiring mid-to-high levels of D/T, while women are more heavily represented in occupations with the highest levels of D/T.

In order to use our 2007 version of O*NET, which is matched to the 2000 Census occupation codes (COC), to study measure trends in caring work over a long time period, it is necessary to approximate a time-consistent set of occupation codes. The 2000 COC codes, used in the CPS during 2003-2010, and the 1990 COC used in the CPS during 1992-2002, were converted back to 1980 COC codes using a probabilistic mapping provided by Census.[21] Beginning with the 2007 O*NET values matched to the 2000 COC codes for all wage and salary workers using the 2006-2008 CPS, we then recalculated each O*NET attribute value for the large 2006-2008 sample based on their 1980 COC codes. Once we had O*NET values based on the 1980 COC codes, we then calculated the means of our O*NET measures of caring and D/T for all wage and salary workers ages 16 and over from 1983 through 2010. We find no discontinuity in our series associated with occupational code breakpoints (1991 vs. 1992 and 2002 vs. 2003), thus enhancing our confidence in the series.

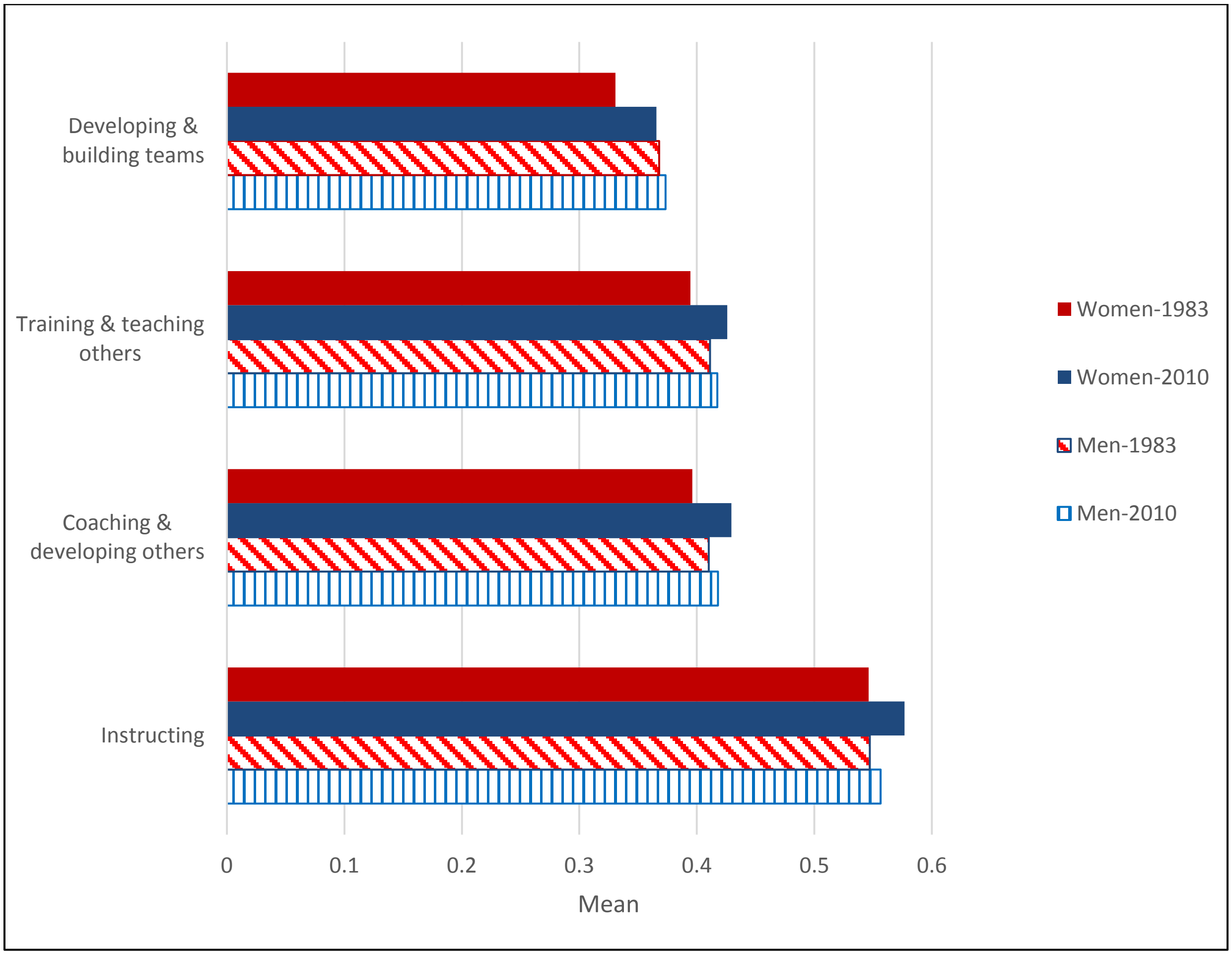
[21] We thank David Macpherson for providing programming code to convert 1990 and 2000 Census occupation codes back to 1980 codes. Note that the O*NET measures matched to 1980 COC codes are used exclusively for the historical series shown in Figs. 3 and 4, but nowhere else in this paper.

Figure 3. Trends in Means of O*NET 'Caring' Variables by Gender, 1983–2010



'A&C' and 'Concern' denote O*NET caring attributes 'assisting & caring for others' and 'concern for others' measured in levels and scaled [0,1]. Weighted means are calculated using all wage & salary workers, ages 16+. Caring measures are from the 2007 O*NET version 12, matched to the 2006-2008 CPS using 2000 COC codes, recalculated for that sample using 1980 COC codes, and then matched to workers for 1983-2010 using probabilistic time-consistent 1980 COC codes. The O*NET values by occupation are fixed. Changes over time are determined by changes in occupational employment. The 'A&C' and 'Concern' measures linked to 1980 COC codes are used in Figures 3 and 4, but nowhere else in the paper. See text for details.

Figure 4. Means of Developing/Teaching O*NET Variables, by Sex for 1983 and 2010



Weighted means are calculated using the full CPS sample for each year excluding only workers under age 16. CPS data are matched based on 1980 Census occupation codes and 2007 O*NET (version 12.0) values. O*NET variables are measured in levels and scaled on [0,1]. See text for further discussion.

The matched CPS/O*NET data set allows us to examine changes in caring in the labor market over the 28 year period from 1983 to 2010. Note that the within-occupation O*NET ratings are not changing (they are fixed at 2007 values). Thus, the measured economy-wide changes in caring occur entirely from changes in the distribution of workers across occupations (all index values are employment weighted using CPS sample weights). Figure 3 shows the trends for the years 1983-2010 in the mean levels of "assisting& caring for others" and "concern for others" among both women and men. As evident in the figure, the levels of these caring tasks in the U.S. labor market has increased steadily over time for both women and men, but the overall change has been quite small and there is no evidence of a narrowing in the caring gap between women and men.

Similar descriptive data for 1983-2010 was also constructed for the four O*NET D/T attributes. In Figure 4, we show the means of each of these attributes for the years 1983 and 2010. In contrast to the evidence on caring, we find that the developing and teaching content of both women and men's jobs have increased over time, but substantially more so for women than men. By 2010, women's jobs involved slightly higher levels than did men's jobs of training and teaching, coaching and developing, and instructing others, and slightly lower levels of developing and building teams.

Measuring Wage Differentials For Caring Work: Methods

Our empirical approach is straightforward. We estimate standard Mincerian semi-log earnings functions in which hourly earnings is a function of accumulated human capital (net of depreciation), proxied by time spent in schooling and the labor market. The human capital earnings function is augmented by inclusion of selected demographic controls, measures of location, and job attributes associated with wage differentials, including various measures of caring. We provide estimates from separate female and male earnings functions. This permits a straightforward way to examine how caring job attributes differently affect the wages of women and men, as emphasized in this literature. We provide estimates using panel as well as cross-section data, thus accounting for worker heterogeneity. And our large sample allows us to examine whether caring wage differentials differ across the private and public sectors, as well as among various worker characteristics.

To examine caring wage differentials across workers, we use the CPS/O*NET 2006-2008 sample and estimate wage level equations of the following form:

$$lnW_{iF} = \sum_{k=1}^{K}(\beta_{kF}X_{ikF}) + \gamma_F caring_{iF} + \varepsilon_{iF} \qquad (1a)$$

$$lnW_{iM} = \sum_{k=1}^{K}(\beta_{kM}X_{ikM}) + \gamma_M caring_{iM} + \varepsilon_{iM} \qquad (1b)$$

Here, subscripts *F* and *M* denote female and male, respectively; lnW_i is the natural log of hourly earnings for worker i; X_{ik} contains an intercept and K − 1 independent variables measuring worker and job-related characteristics; β_k contains a constant and coefficients for covariates in X; $caring_i$ is the covariate(s) of interest measuring caring skills for each worker's detailed occupation; γ are the coefficients of interest measuring marginal wage effects of one standard deviation changes in the caring measures; and ε_i is an idiosyncratic error term. Estimates of γ_F and γ_M may be sensitive to the controls included in X_k, in particular measures of job skills and working conditions. And variants of equations (1a) and (1b) can be estimated within different sectors or for different groups of workers.

If employment in caring jobs is correlated with workers skills, motivation, etc. not reflected in measures of occupation skill requirements, wage level estimates may be biased. In order to account for worker heterogeneity, we use panel analysis, in this case longitudinal wage change equations that conform to the CPS sample structure. To examine wage differentials for caring among job switchers, we use our CPS/O*NET 2003/2004-2008/ 2009 panel sample, including only those workers who change occupations and, thus, have had changes in the O*NET caring measures.[22] We estimate the following wage change equations:

$$\Delta lnW_{iF} = \sum_{k=1}^{K}(\beta_{kF}\Delta X_{ikF}) + \gamma_F \Delta caring_{iF} + \Delta\varepsilon_{iF} \qquad (2a)$$

$$\Delta lnW_{iM} = \sum_{k=1}^{K}(\beta_{kM}\Delta X_{ikM}) + \gamma_M \Delta caring_{iM} + \Delta\varepsilon_{iM} \qquad (2b)$$

where Δ denotes changes between year-pairs (i.e., 2004 minus 2003, etc.). Parameter estimates in equations (2a) and (2b) are based exclusively on occupation switchers and net out worker-specific fixed effects on wages. Two limitations in using the O*NET job skill/task measures warrant mention: (1) the O*NET values matched to each occupation are fixed over time and (2) the value of each O*NET attribute does not vary across workers in a given occupation.[23] We are not concerned with the first issue—relative occupational differences in attributes change gradually and our analysis is for a relatively short time period. The measurement of job attributes at the occupation rather than individual worker level is a more serious concern. There is heterogeneity of job characteristics within detailed occupations and these may differ to some degree by gender as well as across individuals. It is not clear whether or to what extent measurement error in O*NET job attributes is mitigated in the panel analysis

[22] More precisely, we include only those who have changed detailed occupation and industry in order to insure that we include mostly "true" job switchers. Worker descriptions of occupa¬tion are coded by Census employees and can be coded differently one year apart even when there has been no job change. Industry is reported with greater accuracy and restricting the sample to just those who report changes in occupation and industry insures a high probability of true occupational change, thus avoiding attenuation of the caring coefficients. For a careful discussion of this approach, see Macpherson and Hirsch (1995).

[23] O*NET updates ratings for occupations on a rolling basis; it takes several years for all occupations to have revised ratings.

through differencing (i.e., where "two wrongs can make a right"). Because job attributes are at the occupation rather than individual level, we cluster standard errors by occupation in the wage level analysis and by occupation switching pairs in the longitudinal analysis.[24]

An important advantage of our CPS/O*NET data set is that we not only have continuous measures of caring intensity across all occupations, but also measures for a large array of detailed job tasks, skill requirements, and working conditions. Such data make it more likely that we can obtain relatively clean estimates of wage differentials associated with caring in the U.S. labor market. And because the analysis also is done using CPS longitudinal data, where we identify person-specific wage changes resulting from movement across years into or out of occupations with different levels of caring, we can account for bias due to worker heterogeneity correlated with caring.

We include additional independent variables to control for other important factors that may influence wages. When estimating wage level regression equations (1a) and (1b), we use controls for potential experience (years since schooling completed or since age 16, whichever is less) in quartic form, dummies for gender, race, ethnicity, marital status, foreign-born, union, region (8 dummies for 9 regions), city size (6), year, broad industry (11), and broad occupation (9). Education dummies are included for the completed grades 9, 10, 11, and 12 (but no diploma), plus high school degree (including the GED), some college no degree, associate, bachelor, masters, professional, and doctorate degrees (those reporting 0-8 years are the

[24] Because there are such a large number of occupation-to-occupation combinations in the panel analysis, there are only trivial differences between non-clustered and clustered standard errors. An issue not addressed for the wage level analysis is that two observations exist for a sizable share of the workers, thus decreasing standard errors and slightly exaggerating signifi¬cance levels. This is a standard issue when using full CPS samples for adjacent years. For those concerned, the typical approach is to cut the sample (roughly) in half using workers observed either in their first year (rotation group 4) or second year (rotation group 8). We know of no examples where results meaningfully differ between such samples.

omitted category). Dummies are also included for the public sector and the private-not-for-profit sector (private-for-profit being the reference group). Given that caring jobs are sometimes largely female jobs, we separately estimate wage equations including a sex composition variable to control for the ratio of the number of females to total workers in each occupation. Although this is not the focus of our study, we can observe the sensitivity of caring penalty estimates to the inclusion of gender composition (and vice versa).

Caring Wage Differentials Using Wage Level Analysis

Table 5 provides wage level results for women and men, respectively, with different columns (specifications) including alternative measures or combinations of caring. Column 1 includes the single “caring” index that loads the two O*NET descriptors measuring “assisting & caring” and “concern” for others. Column 2 includes the “developing/teaching” index that loads the four relevant O*NET attributes; it excludes the caring index. Column 3 includes both the “caring” and “D/T” factor indices. Column 4 regressions include the six separate O*NET measures (each normalized to zero mean and standard deviation of one) included in the caring and D/T indices, but we focus on (and show) only the O*NET measures “assisting & caring” and “concern” for others (coefficients on the other four are included in appendix Table A1). Coefficients on the O*NET factor indices in columns 1-3 and normalized O*NET descriptors in column 4 can be interpreted as the partial effect of a one standard deviation change in the caring measure.

All specifications include a rich set of individual worker and location controls (see the text note) and, importantly, O*NET occupational skill and working condition indices. For reasons of space, we do not show coefficients for all our control variables. Tables in the

Table 5. Wage Level Regression Estimates for Care Work Effects, by Gender, 2006 – 2008

	Women			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	-0.0431*		-0.0259	
	(0.0223)		(0.0220)	
Develop/Teach Index		-0.0967***	-0.0908***	
		(0.0211)	(0.0229)	
Assisting and Caring for Others				0.0459**
				(0.0227)
Concern for Others				-0.0685***
				(0.0155)
Job Skills Index	0.1939***	0.2638***	0.2683***	0.2683***
	(0.0233)	(0.0342)	(0.0330)	(0.0311)
Working Conditions Index	0.0800**	0.0882**	0.0961***	0.0829***
	(0.0373)	(0.0375)	(0.0356)	(0.0288)
CPS controls	Y	Y	Y	Y
Observations	166,009	166,009	166,009	166,009
R-squared	0.4748	0.4813	0.4820	0.4878
	Men			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	-0.0885***		-0.0732***	
	(0.0163)		(0.0150)	
Develop/Teach Index		-0.0696***	-0.0455***	
		(0.0161)	(0.0123)	
Assisting and Caring for Others				-0.0149
				(0.0172)
Concern for Others				-0.0490***
				(0.0118)
Job Skills Index	0.1775***	0.2039***	0.2094***	0.2134***
	(0.0121)	(0.0177)	(0.0154)	(0.0159)
Working Conditions Index	0.0062	0.0102	0.0132	0.0100
	(0.0142)	(0.0150)	(0.0141)	(0.0144)
CPS controls	Y	Y	Y	Y
Observations	168,760	168,760	168,760	168,760
R-squared	0.4971	0.4945	0.4989	0.4993

Standard errors in parentheses, clustered by occupation. *** p<0.01, ** p<0.05, * p<0.10. Dependent variable is ln(hourly earnings). Regression coefficients measure the wage effects of one standard deviation changes. CPS controls consist of detailed education attainment dummies, demographics (potential experience and its square, cubic, and quartic terms); dummies for marital status, race, ethnicity, foreign-born citizen, non-citizen), geographic dummies for region (8) and MSA size (6), broad industry (11) and broad occupation (9) dummies, year dummies, and dummies for union membership, public sector, and private nonprofit sector. Col (4) includes the four developing/teaching measures. See text for further details.

Table 6. Wage Level Regression Estimates for Care Work Effects by Gender, Absent O*NET Skill & Working Conditions Indices

	Women			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	0.0181		0.0010	
	(0.0349)		(0.0375)	
Develop/Teach Index		0.0404*	0.0401*	
		(0.0209)	(0.0213)	
Assisting and Caring for Others				0.0798*
				(0.0406)
Concern for Others				-0.0849***
				(0.0224)
Job Skills Index	N	N	N	N
Working Conditions Index	N	N	N	N
CPS controls	Y	Y	Y	Y
Observations	166,009	166,009	166,009	166,009
R-squared	0.4415	0.4439	0.4439	0.4598
	Men			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	-0.0313		-0.0630***	
	(0.0203)		(0.0203)	
Develop/Teach Index		0.0333**	0.0564***	
		(0.0162)	(0.0172)	
Assisting and Caring for Others				-0.0096
				(0.0223)
Concern for Others				-0.0457**
				(0.0183)
Job Skills Index	N	N	N	N
Working Conditions Index	N	N	N	N
CPS controls	Y	Y	Y	Y
Observations	168,760	168,760	168,760	168,760
R-squared	0.4686	0.4693	0.4726	0.4771

Standard errors in parentheses, clustered by occupation. *** p<0.01, ** p<0.05, * p<0.10. Dependent variable is ln(hourly earnings). The regressions in this table are identical to those shown in Table 5, except for the exclusion of the Job Skills and Working Condition indices. See Table5 notes.

appendix provide coefficients for most of the control variables not shown in Tables 5 and 7 (the wage level and wage change results, respectively). Table 6 is identical to Table 5, except that we omit the O*NET skill and working condition indices. Before turning to results with respect to the caring variables, we note that the skill (and to a lesser extent, working conditions) indices have large and highly significant coefficients and add 3-4 percentage points to the R^2 values. A one s.d. increase in the skill index is associated with a roughly 0.20 log point increase in wages for women and men. These large wage effects reflect the impact not only from job skill requirements but also from worker skills not fully captured by schooling, potential experience, and other CPS control variables. In our panel analysis (shown subsequently), which accounts for worker fixed effects, the skill index coefficients are only about one-fifth as large, but still highly significant. The estimated working conditions coefficients are positive, as predicted by theory, with larger coefficients (in cross-section analysis) for women than men, consistent with prior literature (Hersch, 1998).

Turning to the estimated effects of caring on wages in Table 5, we first focus on the estimated wage effects of the "caring" index, absent control for the D/T index (column 1). For women, the estimate is a wage penalty of 0.04 log points, which is marginally significant. The coefficient for men is a more substantive and significant coefficient of −0.09, indicating a 9% lower wage with respect to a one s.d. increase in the level of caring. In column 2, we include the D/T but not caring index and in column 3 we include both. The caring coefficients change relatively little with the addition of D/T. The coefficients on D/T are negative and significant for both women and men. Focusing on the individual caring measures in column 4, it is readily evident that for women and men, the "concern" measure is more strongly associated with wage penalties than is the “assisting & caring” measure. Recall that the “concern” measure is intended to measure needed worker attributes, whereas the A&C measure reflects required work activities. A possible interpretation of these results is that the negative concern coefficient captures penalties resulting from labor supply preferences (i.e., the prisoner of love theory).

Overall, the cross-section evidence in Table 5 provides support for the proposition that there are systematic nontrivial caring penalties in the labor market. It is important to note that the caring coefficient estimates are highly sensitive to inclusion of the skill index, consistent with the observation that high caring occupations include some of the highest (and lowest) skilled occupations in the labor market. Table 6 is identical to Table 5, except that each wage equation excludes the O*NET skill and working conditions indices. As evident in Table 6 results, there is no longer clear evidence for caring wage penalties among women or men. Particularly notable is that the D/T coefficients turn from sharply negative to sharply positive once the job skills index is removed. As discussed subsequently, this change is due to the exceptionally high skill requirement ratings attached to teaching occupations. A similar effect occurs with respect to the caring coefficients. Health professional occupations, nursing in particular, have high job skill requirements that we control for in Table 5 but not in Table 6. Because required skills are an important wage determinant and differ across occupations, it is important to control for occupational skills in order to isolate the effects of caring on wages. Absent these controls, the only evidence we find for caring penalties is associated with the O*NET descriptor measuring workers’ need to have concern for others, its coefficients being similar with or without inclusion of the skill index.

Our takeaway points from Tables 5 and 6 are twofold. First, evidence for penalties associated with caring jobs is stronger for men than for women and is rather mixed with respect to the types of caring attached to jobs. Second, skills matter. Worker and job skills have high payoffs in the labor market but are not typically well accounted for (controlled) in standard analyses. Caring jobs sometimes require highly skilled workers and tasks (e.g., registered nurses) and sometimes not. Because of the importance of skill, we place greater emphasis on our longitudinal estimates (shown below), which control both for worker heterogeneity (worker-specific fixed skills) and changes in job skill requirements.

Wage Equation Caring Results Using Longitudinal Analysis

As in England et al. (2002), our preferred approach for estimating caring wage differentials is longitudinal analysis that accounts for worker heterogeneity fixed over time (i.e., consecutive years using the CPS). There are potential downsides to using longitudinal data and, more narrowly, to using a sample that identifies wage gaps exclusively on job switchers (in this application, workers changing both occupation and industry). One concern is whether or not caring wage gaps for the job switcher sample are representative of the larger labor force. A second concern is whether selection into job change (i.e., endogenous job change) biases the longitudinal estimates. We address these issues below.

To examine the representativeness of the job switcher sample, we compare measurable worker characteristics for the two samples and estimate wage level equations for the job switcher sample identical to those shown in Table 5 (we show results using the initial year observation, but results are nearly identical using the second year). In comparing means of all variables for the two samples, we found no large or unusual differences (these results are available on request). The wage level regression results for the job switcher sample are shown in appendix Table A2. The results for the two samples are reasonably similar, although not identical. We see the same pattern of wage level caring penalties for both samples, the principal difference being that the caring penalties are somewhat higher for women and lower for men using the longitudinal rather than full samples. Indeed, the caring coefficients using the job switcher sample are quite similar for women and men, in contrast to those seen in Table 5. Our assessment is that the job switcher sample appears to be roughly representative, based both on similar measured characteristics and similar regression coefficients with comparable models.

The results from the longitudinal wage change analysis, shown in Table 7, are reasonably clear-cut. The magnitude of coefficients is substantially smaller using longitudinal rather than wage level analysis. Specifically, the coefficients on the "caring" index (column1 of Table7) for women indicates a 0.014 log point decrease for a one s.d. increase in the caring index, while the estimate for men shows a 0.018 log point decrease. These coefficients are similar when the D/T index is included (column 3). Coefficients on the D/T index are effectively zero for women and men. In column (4) we break out the caring index into its component parts, obtaining negative but small wage effects for both "caring & assisting" and "concern" for others (the four D/T measures are also included, with coefficients shown in appendix Table A3). These caring estimates are tiny and not significant for women. They are somewhat more substantive for men, 1.1% and 0.8% wage decreases for one s.d. increases in "caring & assisting" and "concern" for others, respectively.

Table 7. Wage Change Estimates for Care Work—Ind/Occ Switchers Only, by Gender, 2003/4 – 2008/09

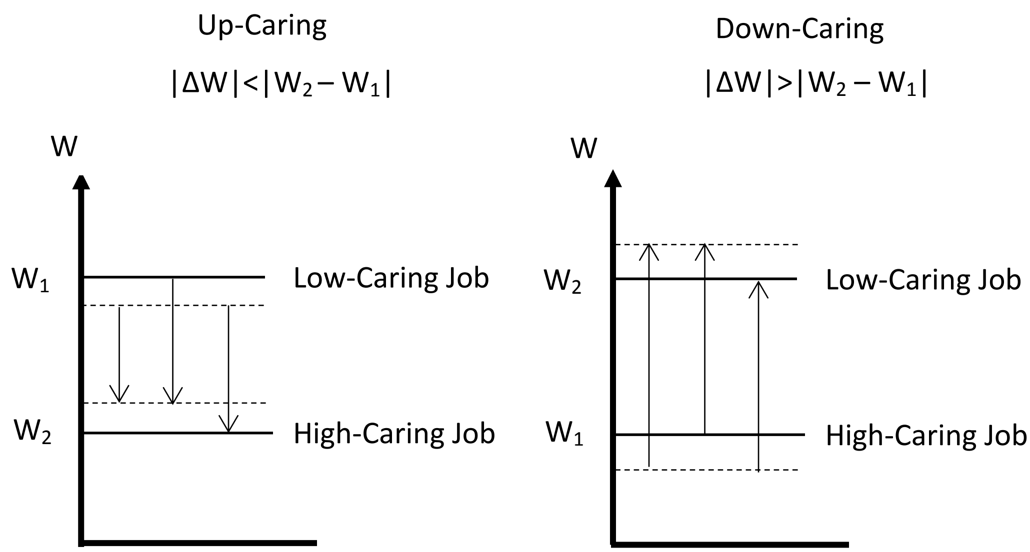
	Women			
Δ O*NET Variables/Indices	(1)	(2)	(3)	(4)
Δ Caring Index	-0.0137***		-0.0140***	
	(0.0043)		(0.0045)	
Δ Develop/Teach Index		-0.0016	0.0017	
		(0.0049)	(0.0051)	
Δ Assisting & caring for others				-0.0058
				(0.0039)
Δ Concern for others				-0.0055
				(0.0040)
Δ Job Skills Index	0.0448***	0.0423***	0.0435***	0.0433***
	(0.0048)	(0.0061)	(0.0062)	(0.0073)
Δ Working Conditions Index	-0.0056	-0.0077	-0.0060	-0.0029
	(0.0060)	(0.0060)	(0.0061)	(0.0061)
Δ CPS controls	Y	Y	Y	Y
Observations	18,981	18,981	18,981	18,981
R-squared	0.0219	0.0213	0.0219	0.0225
	Men			
Δ O*NET Variables/Indices	(1)	(2)	(3)	(4)
Δ Caring Index	-0.0184***		-0.0217***	
	(0.0046)		(0.0047)	
Δ Develop/Teach Index		0.0040	0.0097**	
		(0.0039)	(0.0040)	
Δ Assisting & caring for others				-0.0106**
				(0.0041)
Δ Concern for others				-0.0080**
				(0.0035)
Δ Job Skills Index	0.0407***	0.0331***	0.0346***	0.0302***
	(0.0043)	(0.0048)	(0.0049)	(0.0057)
Δ Working Conditions Index	-0.0009	-0.0021	-0.0021	-0.0020
	(0.0049)	(0.0049)	(0.0049)	(0.0049)
Δ CPS controls	Y	Y	Y	Y
Observations	21,689	21,689	21,689	21,689
R-squared	0.0257	0.0249	0.0259	0.0263

Standard errors in parentheses, clustered by occupation pair. *** p<0.01, ** p<0.05, * p<0.10. Regression coefficients measure the wage effects of one standard deviation changes. ΔCPS controls consist of first differences in higher order terms of experience (square, cubic, and quartic), changes in union membership, public sector, private nonprofit sector, broad industry and broad occupation, and dummies for year pairs. All O*NET variables and indices are differenced. Dependent variable is Δln(hourly earnings).

Overall, we regard the longitudinal evidence of wage penalties for caring work (measured by the narrow caring index) as plausible and reasonably convincing. That said, the magnitude of the estimated penalties is rather small, particularly so as compared to other wage gaps estimated in the larger literature (say, with respect to union status, city size, race/ethnicity, etc.).

A second concern pertinent for the longitudinal analysis is whether selection into job change (i.e., endogenous job change) biases the panel estimates. Selection of course has a bearing on the wage level analysis as well, but there the presumption was that wage differentials observed for caring work are “market prices” determined for marginal workers. Inframarginal workers in caring jobs have a relatively stronger preference for caring and would have accepted lower wages (a largerpenalty) to work in caring jobs. Inframarginal workers in noncaring jobs would not be willing to work in caring jobs at current wage levels. Longitudinal analysis based on occupation switchers is likely to reflect wage changes among workers who tend to be closer to the margin and less likely to be inframarginal, an advantage of such analysis. The concern regarding the panel approach, however, is that even among such a group, job change is endogenous rather than random and likely to be correlated with wage changes. In our panel analysis, endogenous selection is likely to bias caring coefficients in opposite directions (as explained below) depending on whether workers move to occupations with lower or higher levels of caring. Thus, separate estimates for workers who “up-care” and “down-care” provide a useful robustness check on our results.

Figure 5. Bounds for Selection Bias in Estimates of a Wage Penalty for Care Work



W_1 and W_2 are mean market wages for low- and high-caring jobs. The arrows designate the range of wage changes most likely seen among job switchers, and ΔW represents the mean of observed wage changes among the movers. Selection should lead to smaller observed wage changes (in absolute value) among those who “up-care” than among those who “down-care”.

Workers are more likely to switch jobs the less attractive the wage in the current job relative to the wage in the destination job. This can affect measurement of the caring wage differential, as illustrated in Figure 5, which shows workers switching from job 1 with market wage W1 to job 2 with market wage W2. Job switchers who “down-care” (think of this as moving to a job disamenity or, equivalently, a loss in amenities) are more likely to do so if they have a relatively low wage draw on the current job (seen by the dotted line below market wage W1) or high wage draw on the destination job (seen by the dotted line above market wage W2). This leads to an upward bias in the caring wage gap seen among down-care movers in that the observed ΔW exceeds the market wage differential W2− W1. The same logic holds for “up-care” job switchers, who on average are more likely to have made the move if the wage on the lower-caring initial job was less than W1 and/or there is a low wage penalty associated with the higher-caring destination job (i.e., a wage above W2), thus leading to a downward bias in wage gaps seen between the low caring source job and high caring destination job. Imagine that the market equilibrium for a given difference in job caring is a 2% caring wage penalty. Our expectation is that estimates from those who “down-care” would exhibit average wage increases greater than 2%, while those who “up-care” will exhibit wage changes less than 2% in absolute value (i.e., wage decreases less than 2%). If we were to observe the predicted pattern with wide bounds around the average, it would suggest that such selection bias is substantive.

As seen in Table 8, longitudinal wage change equations providing separate estimates for those who up-care and down-care display fairly tight bounds, with similar coefficients for those who up-care and down-care (in no case are coefficients significantly different). For men, the

Table 8. Separate Wage Change Estimates for Care Work for Up-Caring and Down-Caring Ind/Occ Switchers, by Gender, 2003/4 – 2008/09

	Women		
Δ Independent Variables	(1)	(2)	(3)
Up·ΔCaring index	-0.0158**		-0.0166**
	(0.0074)		(0.0075)
Down·ΔCaring index	-0.0115		-0.0114
	(0.0074)		(0.0075)
Up·ΔDevelop/Teach index		0.0008	0.0044
		(0.0067)	(0.0068)
Down·ΔDevelop/Teach index		-0.0040	-0.0011
		(0.0071)	(0.0072)
Δ Job Skills Index	0.0448***	0.0423***	0.0435***
	(0.0048)	(0.0061)	(0.0062)
Δ Working Conditions Index	-0.0057	-0.0078	-0.0060
	(0.0060)	(0.0060)	(0.0061)
Δ CPS controls	Y	Y	Y
Observations	18,981	18,981	18,981
R-squared	0.0220	0.0214	0.0220
	Men		
	(1)	(2)	(3)
Up·ΔCaring index	-0.0129		-0.0135
	(0.0081)		(0.0082)
Down·ΔCaring index	-0.0240***		-0.0299***
	(0.0087)		(0.0088)
Up·ΔDevelop/Teach index		-0.0035	0.0007
		(0.0070)	(0.0071)
Down·ΔDevelop/Teach index		0.0114*	0.0186***
		(0.0068)	(0.0070)
Δ Job Skills Index	0.0407***	0.0332***	0.0347***
	(0.0043)	(0.0048)	(0.0049)
Δ Working Conditions Index	-0.0008	-0.0019	-0.0019
	(0.0049)	(0.0049)	(0.0049)
Δ CPS controls	Y	Y	Y
Observations	21,689	21,689	21,689
R-squared	0.0262	0.0255	0.0266

Standard errors in parentheses, clustered by occupation pair. *** p<0.01, ** p<0.05, * p<0.10. No differences between Up and Down Care coefficients are significant at standard levels. Regression coefficients measure the wage effects of one standard deviation changes. See Table 6 for ΔCPS controls; see text for a detailed discussion of variables including interaction terms.

separate coefficient estimates for "Up ΔCare" and "Down ΔCare" generally produce lower caring wage gap estimates (in absolute value) for workers who up-care and higher estimates for those who down-care, as expected based on selection. This pattern is not generally evident among women, suggesting any bias from selection is small, being more than offset by sample differences in wage changes among the "up" and "down" groups. The important point is that in all cases, coefficient differences between those moving toward more-and less-caring occupations are small, providing relatively tight bounds on our results. As an example, recall that in Table 7, column 3, we found statistically significant penalties of −0.014 for women and −0.022 for men. Estimating separate "up" and "down" care wage changes in column 3 of Table 8, we obtain estimates for men of −0.0135 and −0.030, symmetrically bounding the earlier −0.022 estimate and supporting the expected pattern of selection into job change. For women, we obtained "up" and "down" estimates of −0.017 and −0.011, which provides narrow bounds on the joint −0.014 estimate, but do not show the expected pattern of selection.[25]

Caring Penalties In The Public Versus Private Sectors

Authors emphasizing the importance of caring wage penalties have stressed not only its disproportionate impact on women, but also that it may reflect a concentration of caring jobs in the public sector (e.g., Barron & West, 2013). If caring jobs are concentrated in the public sector, a caring penalty would mechanically arise if public workers are systematically paid less than are similar private sector workers in similarly demanding jobs. Analysis of public-private

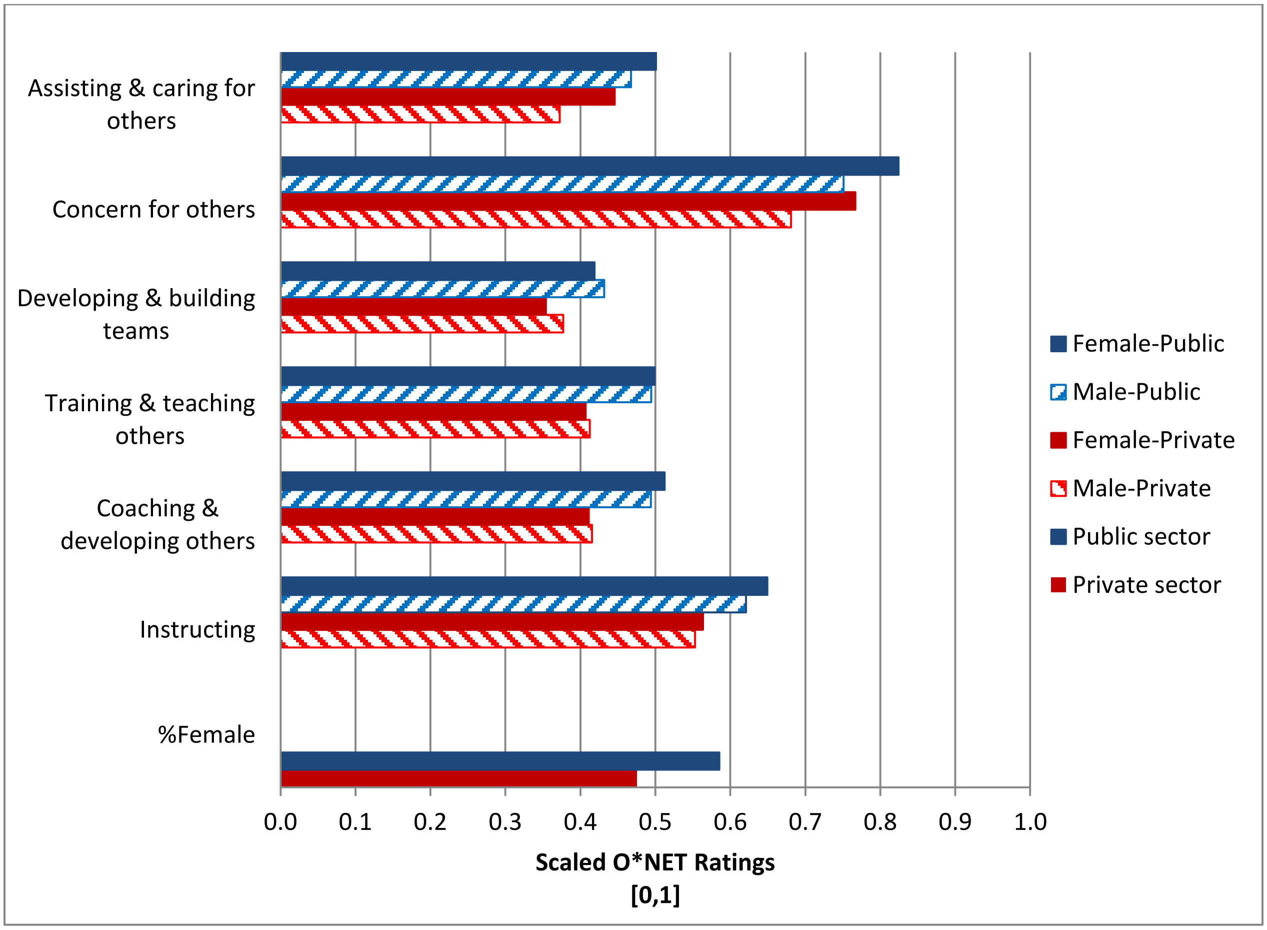
[25] For both women and men and for all measures of caring, there were roughly equal num¬bers of workers "up-caring" and "down-caring" but with the number in the former group slightly exceeding that in the latter, consistent with the gradual upward trend in caring over time (see Fig. 3). For both female and male job switchers, mean changes in caring magnitudes are nearly identical (in absolute value) for "up" versus "down" changes. For women, mean changes in the caring and developing/teaching indices are 0.67 and 0.86 standard deviations. For men, the equivalent means are 0.59 and 0.79.

pay differences is beyond the scope of our paper, but our assessment of a now extensive literature is that estimated public/private wage differentials are typically small, sometimes negative and sometimes positive, and somewhat sensitive to inclusion of what have become controversial control variables (e.g., union status and employer size). More clear-cut is the finding of considerably greater pay compression and relatively higher nonwage benefits (pensions, health insurance) in the public sector.[26] Although not the focus of our study, in our regression analyses, a public sector dummy variable is systematically positive for women and men in the wage level analyses and effectively zero in the wage change analyses (shown in the appendix Tables A1 and A3).

We do not focus on public/private pay differences, but instead examine how wages vary with respect to caring measures within the public and private sectors. Our principal focus is on results from wage level analyses for the two sectors. We do not examine wage changes among those switching occupations within the public sector because of measurement issues (discussed below). In addition to distinguishing between the public and private sectors, caring wage effects can differ across the private for-profit and private not-for-profit sectors, the latter including numerous health care workers employed by not-for-profit hospitals, as well as other caring workers employed by nonprofits other than hospitals. As noted by a referee, public programs (e.g., Medicare and Medicaid) also can influence wage levels in the private sector. To preview our results, we find evidence consistent with caring penalties being more likely in the public than in the private sector, and (seen subsequently in Section 10) with similar wages for caring work in the private for-profit and not-for-profit sectors.

[26] Recent studies include Gittleman and Pierce (2011), Bender and Heywood (2012), and Lewin, Keefe, and Kochan (2012). The debate regarding inclusion of a union control can be traced back to Linneman and Wachter (1990).

Figure 6. O*NET Ratings of Caring Attributes by Gender and Public/Private Sectors



Figures reflect the means shown in Table 9, which are based on the CPS/O*NET 2006-2008 sample. The %Female for public is 58.5% and for private 47.6% in our 2006-2008 sample.

Table 9. O*NET Caring Attribute/Index Means in the Private and Public Sectors, by Gender

	Women		Men		
Private Sector Means:	mean	s.d.	mean	s.d.	$\bar{X}_F - \bar{X}_M$
Hourly earnings	18.22	13.66	22.75	17.96	-4.53
Log hourly earnings	2.7237	0.5648	2.9240	0.6022	-0.2003
Individual O*NET attributes (scaled 0-1):					
Concern for others	0.7668	0.1105	0.6809	0.1045	0.0859
Assisting & caring for others	0.4456	0.1657	0.3725	0.1156	0.0731
Developing & building teams	0.3540	0.1283	0.3772	0.1258	-0.0232
Training & teaching others	0.4069	0.1218	0.4125	0.1119	-0.0056
Coaching & developing others	0.4111	0.1392	0.4156	0.1388	-0.0045
Instructing	0.5636	0.1135	0.5531	0.1001	0.0105
O*NET Indices (using factor analysis):					
Caring index (loads concern and asst/caring above)	0.2097	0.8753	-0.3476	0.6769	0.5573
Developing/Teaching index (loads remaining 4 attributes)	-0.0790	0.9114	-0.0321	0.8629	-0.0469
N	129,987		143,228		
	Women		Men		
Public Sector Means:	mean	s.d.	mean	s.d.	$\bar{X}_F - \bar{X}_M$
Hourly earnings	21.03	12.74	25.77	15.09	-4.74
Log hourly earnings	2.9070	0.5184	3.1164	0.5093	-0.2094
Individual O*NET attributes (scaled 0-1):					
Concern for others	0.8244	0.1076	0.7512	0.1221	0.0732
Assisting & caring for others	0.5009	0.1424	0.4676	0.1610	0.0333
Developing & building teams	0.4188	0.1330	0.4316	0.1309	-0.0128
Training & teaching others	0.4991	0.1573	0.4943	0.1490	0.0048
Coaching & developing others	0.5125	0.1644	0.4941	0.1600	0.0184
Instructing	0.6493	0.1414	0.6209	0.1345	0.0284
O*NET Indices (using factor analysis):					
Caring index (loads concern and asst/caring above)	0.6029	0.8010	0.2172	0.8934	0.3857
Developing/Teaching index (loads remaining 4 attributes)	0.6180	1.1351	0.5374	1.0692	0.0806
N	36,022		25,532		

All means are calculated from the merged 2006-2008 merged CPS/O*NET data set described in the text.

Prior to turning to the wage evidence, we first examine the extent to which caring work is more prevalent in the public than in the private sector based on our multiple O*NET measures of caring. Figure 6 (as well as Table 9) shows the relative values for the six O*NET attribute measures. There is clear-cut evidence that for women and men, jobs in the public sector require considerably higher levels of caring, defined broadly or narrowly. Focusing on the O*NET factor indices (with mean 0 and s.d. 1 over the entire male/ female, private/public sample), each is higher for women than men within the public sector, and each is higher in the public than in the private sector for women and for men. For example, the caring index for women in the public sector is 0.60 versus 0.21 in the private sector; for men, the public and private values are 0.22 and −0.35. For the D/T factor index, the value for women in the public sector is 0.62 versus −0.08 in the private sector; for men, the public and private values are 0.54 and −0.03. And as widely recognized, women are disproportionately employed in public sector jobs. The share of women in our public sector sample is 58.5%, as compared to 47.6% in the private sector. Using all CPS wage and salary workers for 2007 (i.e., no sample exclusions) and employing sample weights, the shares of women in the public and private sectors are 57% and 46%, respectively.[27]

Turning to the wage level results (Table 10), we provide separate estimates for caring wage penalties within the private and public sectors. In the private sector, results are similar to those seen previously in Table 5 for the full sample, where roughly five out of six workers are in

[27] Men are slightly underrepresented in our estimation samples because they have higher rates of earnings non-response. As discussed previously, non-respondents are not matched to donors based either on public/private status or detailed occupation;their inclusion would attenuate esti¬mated wage gaps(Bollinger&Hirsch, 2006).

the private sector. As before, evidence for caring penalties among women is weak; that for men is more clear-cut.

For the public sector (second page of Table 10), estimates for women and men suggest substantive caring penalties. Among women, a 5.7% penalty (column 3) is associated with a one s.d. change in the caring index. Among men, a 9.2% penalty (column 3) is found. For women and men, caring penalties in the public sector appear to be driven by jobs requiring “concern” and not “assisting & caring” for others (column 4). As noted previously, the “concern” variable is intended to measure needed worker attributes, whereas the “assisting& caring” variable measures job activities.

Notable in Table 10 are also the significantly negative coefficients attached to the D/T index, a result highly influenced by the large number of teachers in the sample. Our assessment is that these apparent penalties are overstated due to exceptionally high O*NET skill requirement ratings assigned to teaching occupations.[28]

Our wage level analysis in this section supports the thesis of larger caring penalties in the public than in the private sector. Because of unobserved worker heterogeneity, we have attached greater weight to the longitudinal than to the cross-section results. Economy-wide,

[28] The issue here is the impact of our occupational skills index. Note how the D/T index coefficients for our full sample turn from negative to positive when the skill index is excluded (Table 5 versus 6). O*NET ratings of the skills involved in teaching are extremely high. Given the large market rewards associated with the job skill index, one finds that teachers are under¬paid in the sense that their hourly earnings fall below the predicted wage. O*NET skill mea¬sures reflect the skills and tasks needed to perform a job well. They do not necessarily provide an accurate measure of the skills of workers hired in these jobs, although competitive market forces should limit discrepancies between worker skills and job skill ratings. It is quite possible that O*NET skill ratings for teaching jobs overstate the skill level of the average teacher. Thus, wage analyses controlling for job skill requirements understate relative pay for teachers. For example, Allegretto, Corcoran, and Mishel (2004) use the CPS to compare hourly earnings for teachers with non-teachers, controlling for CPS worker attributes plus an occupational work level index derived from BLS data, similar in spirit to our O*NET skill index. Teachers were rated very high in required job skills and the authors concluded that teachers are substantially underpaid. Other analyses in the literature (e.g., Podgursky, Monroe, & Watson, 2004; Scafidi, Sjoquist, & Stinebrickner, 2006) fail to support the thesis of teacher underpayment, finding that those who leave teaching suffer substantive wage losses.

Table 10. Wage Level Estimates for Care Work in Public and Private Sectors, by Gender, 2006 – 2008

	Women in the private sector			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	-0.0181		-0.0090	
	(0.0225)		(0.0229)	
Develop/Teach Index		-0.0700***	-0.0685***	
		(0.0207)	(0.0219)	
Assisting and Caring for Others				0.0479*
				(0.0252)
Concern for Others				-0.0546***
				(0.0175)
Job Skills Index	0.1909***	0.2452***	0.2474***	0.2528***
	(0.0228)	(0.0344)	(0.0318)	(0.0331)
Working Conditions Index	0.0509	0.0620	0.0653*	0.0604**
	(0.0340)	(0.0391)	(0.0342)	(0.0290)
CPS controls	Y	Y	Y	Y
Observations	129,987	129,987	129,987	129,987
R-squared	0.4804	0.4842	0.4842	0.4892
	Men in the private sector			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	-0.0640***		-0.0539***	
	(0.0183)		(0.0171)	
Develop/Teach Index		-0.0487***	-0.0354***	
		(0.0146)	(0.0116)	
Assisting and Caring for Others				-0.0199
				(0.0194)
Concern for Others				-0.0289***
				(0.0108)
Job Skills Index	0.1706***	0.1921***	0.1966***	0.1944***
	(0.0141)	(0.0183)	(0.0166)	(0.0170)
Working Conditions Index	-0.0067	-0.0018	-0.0005	-0.0009
	(0.0144)	(0.0145)	(0.0143)	(0.0143)
CPS controls	Y	Y	Y	Y
Observations	143,228	143,228	143,228	143,228
R-squared	0.5066	0.5056	0.5076	0.5077

(continued on next page)

Table 10 (continued). Wage Level Estimates for Care Work in Public and Private Sectors, by Gender

	Women in the public sector			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	-0.0844***		-0.0572***	
	(0.0213)		(0.0192)	
Develop/Teach Index		-0.1196***	-0.0970***	
		(0.0190)	(0.0224)	
Assisting and Caring for Others				0.0133
				(0.0192)
Concern for Others				-0.0712***
				(0.0140)
Job Skills Index	0.1996***	0.2804***	0.2799***	0.2582***
	(0.0298)	(0.0370)	(0.0368)	(0.0389)
Working Conditions Index	0.1312***	0.1410***	0.1507***	0.1282***
	(0.0370)	(0.0330)	(0.0332)	(0.0303)
CPS controls	Y	Y	Y	Y
Observations	36,022	36,022	36,022	36,022
R-squared	0.4512	0.4556	0.4599	0.4636
	Men in the public sector			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	-0.1058***		-0.0916***	
	(0.0128)		(0.0131)	
Develop/Teach Index		-0.0912***	-0.0344**	
		(0.0173)	(0.0160)	
Assisting and Caring for Others				-0.0113
				(0.0149)
Concern for Others				-0.0657***
				(0.0117)
Job Skills Index	0.1882***	0.2222***	0.2122***	0.2052***
	(0.0132)	(0.0189)	(0.0191)	(0.0167)
Working Conditions Index	0.0303**	0.0346*	0.0355**	0.0231
	(0.0153)	(0.0193)	(0.0163)	(0.0143)
CPS controls	Y	Y	Y	Y
Observations	25,532	25,532	25,532	25,532
R-squared	0.4355	0.4245	0.4366	0.4397

Standard errors in parentheses, clustered by occupation. *** p<0.01, ** p<0.05, * p<0.10. Regression coefficients measure the wage effects of one standard deviation changes. CPS controls consist of detailed education attainment dummies, demographics (potential experience and its square, cubic, and quartic terms; dummies for marital status, race, ethnicity, foreign-born citizen, non-citizen), geographic dummies for region (8) and size (6), broad industry (11; private sector only) and broad occupation (9) dummies, and private non-profit (private sector only). Dependent variable is ln(hourly earnings), see text for a detailed description of variables.

the longitudinal evidence suggests much lower caring wage penalties than seen in the wage level analysis (Tables 5 and 7), indicating that worker heterogeneity is important. It is not obvious how to conduct meaningful longitudinal analysis within the public sector, however, because turnover is low and there are relatively few teachers, police, and firefighters (among others) switching both occupation and industry within the public sector. Were we to base such an analysis on changes in recorded occupation codes (but not require an industry change), the ratio of noise (i.e., reporting error) to true signal would be high and estimates would be severely attenuated.

Caring Penalties Across The Earnings Distribution

Prior studies on caring wage differentials have focused on particular occupations, some requiring relatively low and some high skills. We examine how caring wage differentials vary across the earnings distribution using two approaches. First, we provide estimates using quantile regression, which provide caring coefficient estimates that vary across the wage distribution. Second, we divide our sample into quintiles based on their predicted earnings using only nonjob attributes (schooling, demographics, location, etc.), which provides an index of earnings attributes or endowments, each weighted by its importance (coefficient) in the earnings function. We then provide OLS estimates within each of the quintiles. This latter approach can be used for our wage change analysis, where quantile regression is not appropriate.[29]

[29] Quantile regression has been used in numerous applications. For examples and further details see Buchinsky (1998) and Koenker (2005). For an example of OLS log wage and wage change regressions for worker groups ordered by predicted wages, see Card (1996).

In Table 11, we show quantile regression results for women and men that correspond exactly to columns (1) through (3) shown previously in Table 5. These regressions include our full set of covariates, plus the caring index alone, the D/T index alone, and both the caring and D/T indices. We show our previous OLS results in the first column, followed by the quantile regression coefficients for percentiles 10, 25, 50, 75, and 90. The OLS results include clustered standard errors, while unclustered standard errors are provided for the quantile regression results. Given that standard errors in the quantile regression model should be less precise than with OLS, our guesstimate is that quantile regression coefficients need to exceed (in absolute value) roughly 0.05 to be significant at the standard 0.05 level.

Table 11. Quantile Regression Wage Level Estimates for Care Work—by Gender, 2006 – 2008

		Women, N=166,009				
O*NET Caring Indices	OLS/Table 5	QR-p10	QR-p25	QR-p50	QR-p75	QR-p90
Caring Index	-0.0431*	-0.0522	-0.0500	-0.0434	-0.0437	-0.0458
	(0.0223)	(0.0027)	(0.0019)	(0.0018)	(0.0022)	(0.0031)
Develop/Teach Index	-0.0967***	-0.0787	-0.0898	-0.0954	-0.0998	-0.1031
	(0.0211)	(0.0028)	(0.0020)	(0.0020)	(0.0024)	(0.0033)
Caring Index	-0.0259	-0.0372	-0.0348	-0.0284	-0.0289	-0.0296
	(0.0220)	(0.0026)	(0.0018)	(0.0019)	(0.0023)	(0.0032)
Develop/Teach Index	-0.0908***	-0.0693	-0.0818	-0.0890	-0.0944	-0.0965
	(0.0229)	(0.0029)	(0.0020)	(0.0021)	(0.0025)	(0.0035)
		Men, N=168,760				
O*NET Caring Indices	OLS/Table 5	QR-p10	QR-p25	QR-p50	QR-p75	QR-p90
Caring Index	-0.0885***	-0.1182	-0.1084	-0.0964	-0.0722	-0.0473
	(0.0163)	(0.0029)	(0.0022)	(0.0020)	(0.0023)	(0.0034)
Develop/Teach Index	-0.0696***	-0.0637	-0.0692	-0.0690	-0.0646	-0.0588
	(0.0161)	(0.0030)	(0.0022)	(0.0020)	(0.0023)	(0.0033)
Caring Index	-0.0732***	-0.1071	-0.0957	-0.0819	-0.0577	-0.0327
	(0.0150)	(0.0031)	(0.0023)	(0.0021)	(0.0024)	(0.0036)
Develop/Teach Index	-0.0455***	-0.0305	-0.0384	-0.0434	-0.0475	-0.0494
	(0.0123)	(0.0030)	(0.0023)	(0.0021)	(0.0024)	(0.0036)

*** p<0.01, ** p<0.05, * p<0.10 for OLS only. Standard errors for OLS are clustered by occupation. QR standard errors not clustered and QR significance levels not shown. "QR-pN" designates quantile regression results at the N^{th} percentile. See Table 5 notes for independent variables.

The pattern of results in Table 11 is reasonably clear-cut. For women, the caring coefficients are remarkably stable across the wage distribution. The coefficient on the caring index (specification 1) indicates penalties of 0.05, 0.05, 0.04, 0.04, and 0.05 log points as one moves from the 10th to the 90th percentiles. Among men, there is the suggestion that penalties are somewhat higher in the bottom half the distribution, with caring penalties of 0.12, 0.11, 0.10, 0.07, and 0.05 log points as one move up the distribution. For all specifications, the median regression results are similar to the OLS mean regression seen previously in Table 5. In short, the quantile regression analysis for women suggests little difference in the wage-caring gradient across the distribution, while that for men suggests smaller penalties toward the top of the wage distribution. Although our principal focus is the wage effect of caring rather than of developing/teaching tasks in the labor market, the D/T coefficients are remarkably stable throughout the earnings distribution.

In Table 12, we provide a distributional analysis of wage-caring gradients using both wage level and wage change analysis. Longitudinal (wage change) analysis does not lend itself to quantile regression; we are not asking how caring coefficients differ across the distribution of wage changes (i.e., the dependent variable). It is informative, however, to estimate OLS wage change (and wage level) equations within predicted wage percentile ranges, as seen in Card (1996). Here, we estimate OLS wage equations for each of the five quintiles of the predicted log wage, based on non-job attributes. The predicted wage provides a convenient index of compensable worker endowments and location wage differences. In the top half of Table 12,

Table 12. Wage Level and Wage Change Estimates for Care Work by Predicted Wage Quintile

	Wage Level Regression Estimates, 2006-2008					
	Women					
O*NET Caring Indices	Table 5	Q1	Q2	Q3	Q4	Q5
Caring Index	-0.0431*	-0.0301	-0.0472**	-0.0413*	-0.0441**	-0.0430**
	(0.0223)	(0.0183)	(0.0202)	(0.0245)	(0.0215)	(0.0175)
Develop/Teach Index	-0.0967***	-0.0156	-0.0625***	-0.0939***	-0.1138***	-0.1239***
	(0.0211)	(0.0213)	(0.0208)	(0.0216)	(0.0246)	(0.0255)
Caring Index	-0.0259	-0.0310*	-0.0441**	-0.0293	-0.0143	0.0004
	(0.0220)	(0.0184)	(0.0203)	(0.0243)	(0.0223)	(0.0168)
Develop/Teach Index	-0.0908***	-0.0175	-0.0598***	-0.0892***	-0.1092***	-0.1240***
	(0.0229)	(0.0197)	(0.0192)	(0.0221)	(0.0281)	(0.0285)
Observations	166,009	33,202	33,202	33,202	33,202	33,201
	Men					
Caring Index	-0.0885***	-0.0302**	-0.0768***	-0.0818***	-0.0916***	-0.0686***
	(0.0163)	(0.0127)	(0.0143)	(0.0140)	(0.0192)	(0.0192)
Develop/Teach Index	-0.0696***	-0.0140	-0.0341***	-0.0440***	-0.0818***	-0.1019***
	(0.0161)	(0.0127)	(0.0113)	(0.0120)	(0.0177)	(0.0180)
Caring Index	-0.0732***	-0.0288**	-0.0718***	-0.0743***	-0.0719***	-0.0330*
	(0.0150)	(0.0130)	(0.0146)	(0.0140)	(0.0176)	(0.0177)
Develop/Teach Index	-0.0455***	-0.0114	-0.0201**	-0.0243**	-0.0542***	-0.0854***
	(0.0123)	(0.0127)	(0.0101)	(0.0094)	(0.0142)	(0.0192)
Observations	168,760	33,752	33,752	33,752	33,752	33,752
	Wage Change Estimates for Ind/Occ Switchers, 2003/4 - 2008/9					
	Women					
O*NET Caring Indices	Table 7	Q1	Q2	Q3	Q4	Q5
Δ Caring Index	-0.0137***	-0.0148	-0.0066	0.0013	-0.0223***	-0.0171*
	(0.0043)	(0.0096)	(0.0104)	(0.0094)	(0.0086)	(0.0090)
Δ Develop/Teach Index	-0.0016	0.0301***	-0.0026	-0.0095	-0.0020	-0.0125
	(0.0049)	(0.0107)	(0.0104)	(0.0100)	(0.0101)	(0.0103)
Δ Caring Index	-0.0140***	-0.0164*	-0.0064	0.0033	-0.0236***	-0.0148
	(0.0045)	(0.0097)	(0.0104)	(0.0098)	(0.0091)	(0.0096)
Δ Develop/Teach Index	0.0017	0.0310***	-0.0020	-0.0101	0.0047	-0.0063
	(0.0051)	(0.0108)	(0.0105)	(0.0104)	(0.0107)	(0.0110)
Observations	18,981	3,797	3,796	3,796	3,796	3,796
	Men					
Δ Caring Index	-0.0184***	-0.0157	-0.0430***	0.0020	-0.0102	-0.0236**
	(0.0046)	(0.0108)	(0.0105)	(0.0106)	(0.0097)	(0.0100)
Δ Develop/Teach Index	0.0040	0.0057	-0.0129	0.0164*	0.0004	0.0074
	(0.0039)	(0.0083)	(0.0093)	(0.0091)	(0.0087)	(0.0099)
Δ Caring Index	-0.0216***	-0.0168	-0.0416***	-0.0036	-0.0119	-0.0326***
	(0.0047)	(0.0107)	(0.0109)	(0.0108)	(0.0105)	(0.0105)
Δ Develop/Teach Index	0.0096**	0.0074	-0.0051	0.0173*	0.0042	0.0207**
	(0.0040)	(0.0083)	(0.0098)	(0.0094)	(0.0094)	(0.0103)
Observations	21,689	4,338	4,338	4,338	4,338	4,337

Standard errors in parentheses, clustered by occupation in wage level results (top half of table) and by occupation-pair in wage change results (bottom half). *** p<0.01, ** p<0.05, * p<0.10. Wage quintiles are based on predicted log wage from wage level estimates (excluding job attributes) using the 2006-2008 CPS sample in the top half of the table and in the bottom half the initial CPS panel year samples, 2003/4-2008/9. Covariates include demographic, education, and geographic variables, plus year dummies. All estimates are OLS.

we show wage level results for women and men; in the bottom half we show wage change results.

The wage level results shown in the top half of Table 12 are roughly supportive of the pattern found using quantile regression, with OLS results at the middle quintile being very similar to median regression results. As with the quantile regression, we do not see systematic patterns of either increasing or decreasing caring wage penalties as we move across the distribution.

Our principal interest is the bottom half of Table 12, which provides OLS wage change results estimated separately by endowment quintiles; the results from the full sample, as shown previously in Table 7, are shown in the first column. For women, coefficient estimates on the caring index are negative and surprisingly similar in the lowest and highest quintiles (−0.015 and −0.017), while coefficients in the middle quintile are near zero. The caring coefficients for men are −0.016 and −0.024 in the lowest and highest quintiles, with near zero in the middle quintile. Because caring jobs are often concentrated among high and low skill workers, there may be too little variation in caring in the middle of the distribution to precisely estimate the wage-caring gradient.

We did not have strong priors that caring wage penalties should be systematically larger or smaller at different parts of the earnings and skill distributions. The quantile regression and wage quintile analyses presented in this section provide little evidence for either an upward-or downward-sloping wage-caring gradient over these distributions.

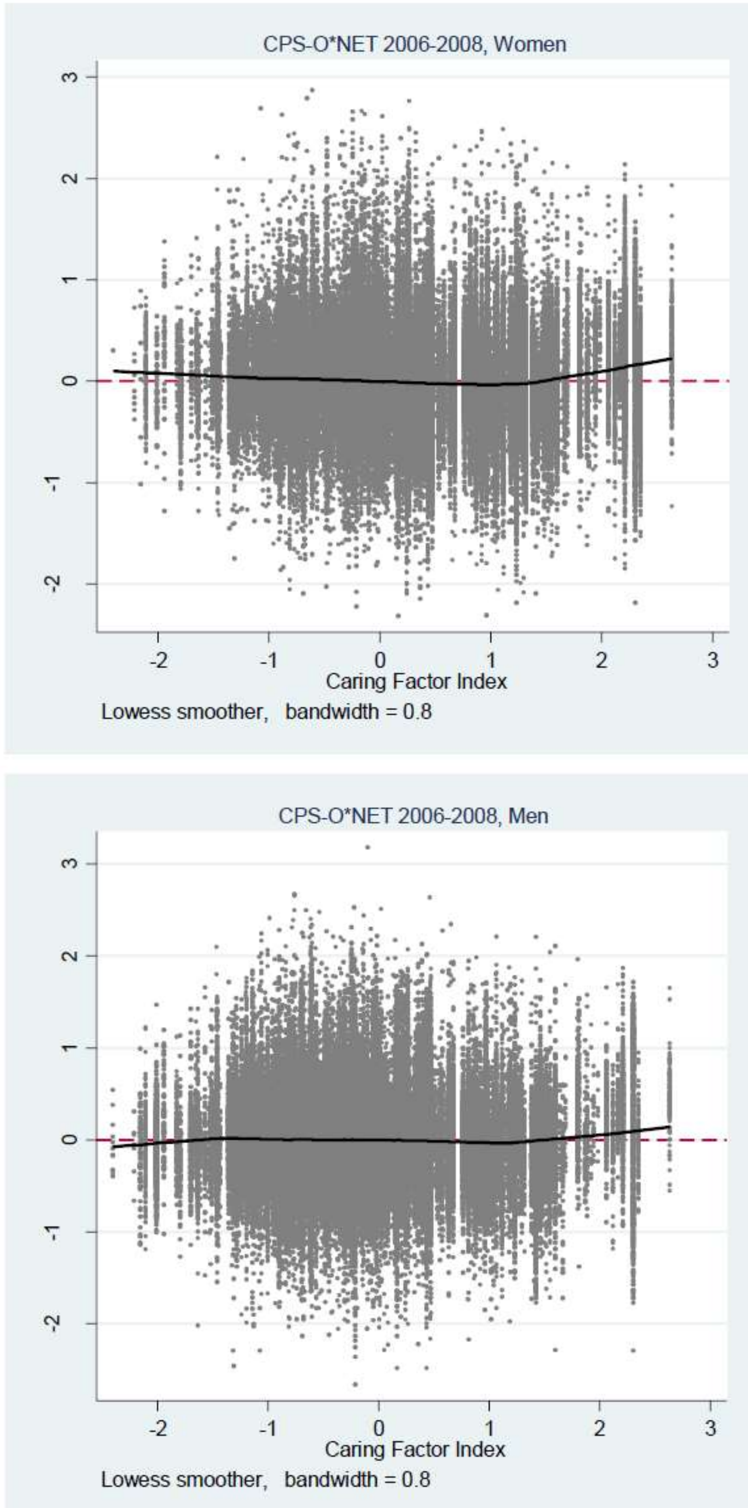
Linearity, Hours Worked, and Group Differences in Caring Wage Effects

In this section, we address three issues. We first examine whether our (implicit) assumption of a linear relationship between wages and the level of caring is reasonable. Second, we ask how weekly hours worked for women and men vary with respect to levels of caring. And third, we examine whether there exist substantive group differences in the relationship between wages and caring.

In order to examine the linearity of the wage-caring gradient, we first extract the residuals from our female and male log wage regressions (specifically, column3 from Table5). We provide a scatterplot of these residuals with respect to the caring index that loads the O*NET measures "assisting & caring" and "concern" for others. By construction the regression residuals have mean zero. If the wage-caring relationship is truly linear, then the plot of mean residuals at each level of caring (in effect, the means for each of 501 occupations ordered by caring level), should be relatively flat at near-zero values. Figure 7 shows this relationship for women and men. The smoothed scatterplots are fitted from locally weighted regressions using the "lowess" command in Stata (Cleveland, 1979).

What is readily evident for both women and men is that there is little slope over most of the distribution and minimal deviation from zero. Absence of substantial deviation from linearity provides support for our using simple and convenient linear specifications. That said the figures are informative. The notable deviation from linearity is with respect to very high caring jobs (heavily dominated by health care jobs), with the upward slope more evident for women than men. The high-caring jobs are dominated by health care, all occupations with a caring index level above 2.0 being in health care, and those between 1.5 and 2.0 being mostly in

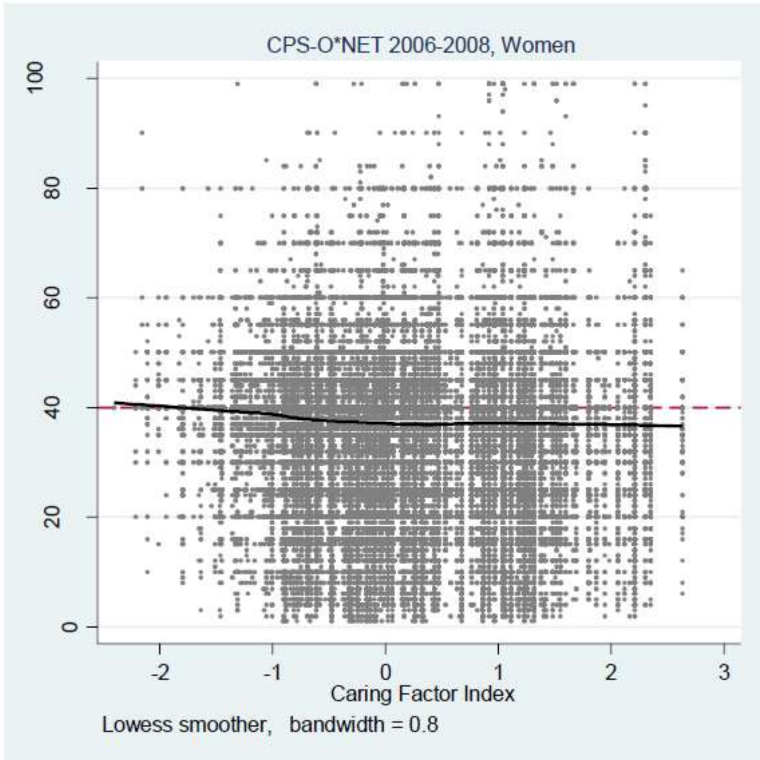
Figure 7. Scatterplot of Wage Residuals and Occupational Levels of Caring for Women and Men

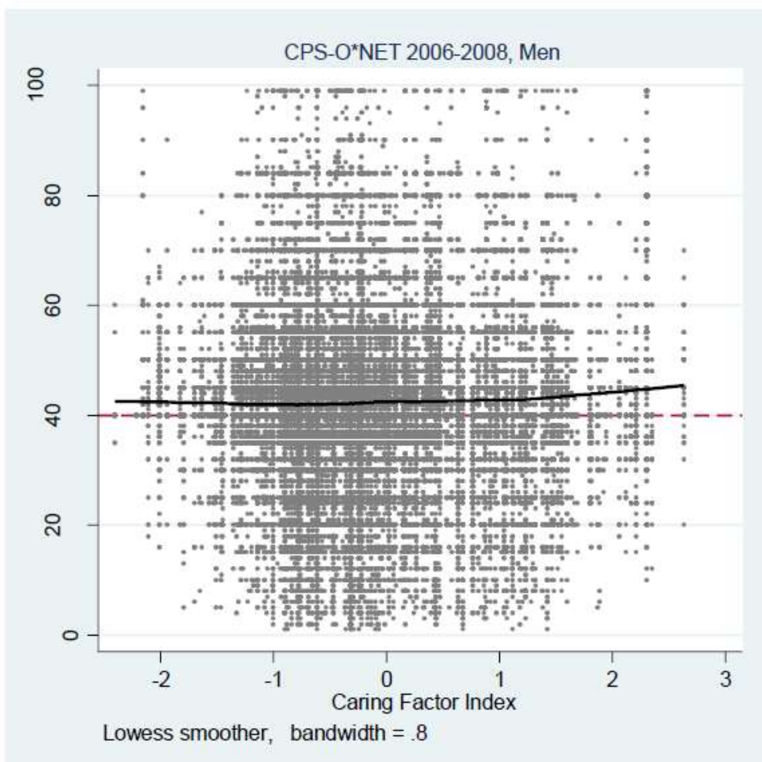


health care. All else the same, workers in health care occupations are paid more than those outside the health sector. While there exist caring wage penalties in some sectors of the labor market, we find little evidence for such penalties in healthcare, the sector of the economy where jobs are evaluated as requiring the highest levels of assisting/caring and, to a somewhat lesser extent, concern for others.

Next we examine how hours worked vary with respect to caring work. Throughout the paper, we examine how average hourly earnings (i.e., wages) differ with respect to caring. If high-caring jobs had substantially higher (lower) weekly hours of work than did low-caring jobs, a substantive caring wage penalty would be associated with even larger (smaller) differences in weekly earnings. To examine the hours-caring relationship, Figure 8 provides a smoothed scatterplot (using the lowess procedure) showing the relationship between usual weekly hours and our O*NET caring index for women and men. Mean hours worked per week are 37.2 hours for women and 42.2 for men (Table2). As seen in Figure 8, in some of the lowest-caring occupations mean weekly hours are close to 40 for women, but then decline with respect to caring over the lower and mid-caring portions of the distribution, remaining roughly constant at well under 40 hours per week in high-caring occupations. In contrast, men's average hours worked are relatively flat over the lower half of the caring distribution, but then turn up and are highest among those working in occupations with the highest levels of caring. Recall that health care occupations dominate the right tail of the caring distribution. Women are overrepresented in health care occupations where part-time work or low full-time hours are common; men are highly represented in health care occupations where long work hours are common.

Figure 8. Scatterplot of Hours Worked and Occupational Levels of Caring for Women and Men





Up to this point, we have estimated the wage-caring relationship separately for women and men, but have not examined differences across other groups of workers. Although such differences are not a focus of the paper, in Table 13 we provide caring level means and wage regression coefficients for selected demographic and worker groups. The caring coefficient for each group is from a wage level regression equivalent to that in column 3 of Table 5.[30] Among both women and men, higher levels of caring are seen for part-time than full-time workers, for salaried versus hourly workers, for non-Hispanics than for Hispanics, for citizens than for noncitizens, among prime-age and older workers than among young workers, in the private not-for-profit and public sectors versus the private for-profit sector, and, most notably, among those with higher education levels.

Estimates of caring wage penalties (i.e., coefficients on the caring index) differ across groups, but often by little. Consistent with previous evidence for women, caring coefficients for most female groups are generally negative, but small and often insignificant. In contrast, all caring coefficients for men are negative, with most being statistically significant and substantive. Among the patterns observed are a tendency for more sizable caring penalties for full-time than for part-time work, increasing caring penalties with age, particularly large penalties for high school graduates and little evidence for penalties among those with graduate and professional degrees, and, as reported previously, more substantive penalties in the public than private sector, but with little difference between the private for-profit and not-for-profit sectors. Few differences are found with respect to race, ethnicity, foreign-born status, or

[30] This specification includes both the caring index and the developing/teaching (D/T) index. The choice of regressors is slightly modified as appropriate for each group. For example, the part-time dummy is excluded when we provide separate estimates for full-and part-time workers

Table 13. Wage Level Estimates for Care Work Effects for Selected Groups, 2006 – 2008

Group	Women, N=166,009			Men, N=168,760		
	N	mean caring	caring estimate	N	mean caring	caring estimate
Full-time	130,924	0.2690	-0.0383*	157,672	-0.2696	-0.0741***
Part-time	35,085	0.3920	0.0284	11,088	-0.1566	-0.0364*
Hourly workers	99,951	0.2480	-0.0015	93,086	-0.3760	-0.0489***
Salaried workers	66,058	0.3662	-0.0392**	75,674	-0.1222	-0.0647***
Non-Hispanic white	123,125	0.3333	-0.0193	123,963	-0.2402	-0.0697***
Non-Hispanic Black	15,199	0.2944	-0.0484***	11,101	-0.1778	-0.0896***
Non-Hispanic other	10,519	0.2087	-0.0281	10,625	-0.2723	-0.0746***
Hispanic	17,166	0.0738	-0.0469**	23,071	-0.4162	-0.0578***
Native citizen	147,637	0.3232	-0.0245	143,851	-0.2354	-0.0711***
Foreign-born citizen	8,208	0.2190	-0.0345	8,207	-0.2889	-0.0666***
Foreign-born non-citizen	10,164	-0.0527	-0.0290	16,702	-0.4793	-0.0728***
Ages 18-24	15,638	0.1684	-0.0133	17,230	-0.3925	-0.0419***
Ages 25-34	38,197	0.3089	-0.0146	41,327	-0.2804	-0.0560***
Ages 35-44	41,823	0.2906	-0.0234	43,532	-0.2469	-0.0555***
Ages 45-54	44,166	0.3076	-0.0389	41,707	-0.2511	-0.0764***
Ages 55-65	26,185	0.3361	-0.0286	24,964	-0.1870	-0.0929***
Never married	38,147	0.2125	-0.0232	43,869	-0.3124	-0.0662***
Ever-married	127,862	0.3196	-0.0278	124,891	-0.2445	-0.0738***
HS dropout	10,745	-0.1642	-0.0476**	17,840	-0.5394	-0.0332***
HS graduate –no college	45,311	0.0452	-0.0653***	51,936	-0.4125	-0.0667***
Some college	52,555	0.3351	-0.0228	46,348	-0.2507	-0.0607***
BA/BS degree	38,142	0.4614	-0.0023	34,787	-0.1369	-0.0614***
Graduate/professional/doctoral degree	19,256	0.7003	0.0122	17,849	0.1784	-0.0131
Large metro (pop ≥ million)	20,426	0.2781	-0.0449*	21,258	-0.2464	-0.0818***
Small metro (100K ≤ pop < 5million)	99,490	0.2862	-0.0247	101,266	-0.2601	-0.0745***
CPS non-metro	46,093	0.3215	-0.0095	46,236	-0.2739	-0.0552***
Private for-profit	111,600	0.1088	-0.0138	135,439	-0.3994	-0.0396***
Private not-for-profit	18,387	0.8221	0.0097	7,789	0.5527	-0.0432*
Public	36,022	0.6029	-0.0572***	25,532	0.2172	-0.0916***

*** p<0.01, ** p<0.05, and * p<0.01 using standard errors clustered on occupation. “Mean caring” is mean of the O*NET caring index. “Caring estimate” is the wage regression coefficient on the caring index. Regression model used is similar to Table 5 Column 3. See Table 5 notes.

marital status. The positive caring coefficients among women for part-time and private not-for-profit are driven primarily by the high wages seen for registered nurses. There is little if any part-time penalty among RNs. Moreover, RNs employed in hospitals earn substantially more than do other RNs (Schumacher& Hirsch, 1997) and about a third of hospital employees (using the 2008 CPS) work for private not-for-profit firms (Hirsch & Macpherson, 2014, Table 7b, p. 97). More broadly, with RNs removed from all the analyses in this paper, one systematically finds more substantial evidence for caring wage penalties among women.

Previous literature has included dummy variables for selected occupations deemed as caring occupations. In practice, this approach leads to designating as caring jobs those who work as teachers, in health care occupations, and in child care and other caring occupations. Our concern about such an approach is twofold. First, such an approach fails to make distinctions between the level of caring required among occupations designated as caring and among those not designated as caring. Second, occupations designated as caring, say registered nurses and teachers, may require different types of caring and entail different types of job tasks. A virtue of our approach is that we use multiple measures -assisting and caring, concern for others, and developing/teaching tasks -each measured as continuous variables for all occupations. Such an approach allows one to distinguish between types of caring, to measure required levels of each task for all jobs, and to differentiate how these tasks are compensated in the labor market, conditional on other wage determinants.

We examine results using the dummy variable approach. Using wage level analysis and a single dummy variable for caring jobs (including health, teaching, and other caring occupations), we obtain negative and significant coefficients for women (−0.11) and men

(−0.26). These results mask what are large positive and negative coefficients found when we break up the caring jobs into medical (2 categories), education (3), and other caring (2). For both women and men, we find large positive coefficients for the health care occupation dummies and large negative coefficients for the education and other caring dummies. When we shift to longitudinal analysis (similar to that in England et al., 2002), coefficients on the caring dummy change variable are virtually zero (−.0006 for women and −.0095 for men), neither being close to statistical significance. These results reinforce our earlier conclusion that accounting for individual worker heterogeneity is important.

Caring Penalties, Occupational Gender Composition, and the Gender Wage

It is widely recognized that both women and men who work in occupations with high proportions of female workers tend to have lower wages (for comprehensive treatments, see Bayard, Hellerstein, Neumark, & Troske, 2003; Ludsteck, 2014; Macpherson & Hirsch, 1995). Less clear are the reasons for these relationships, with it being some combination of sex-based discrimination, differences in job skill requirements and working conditions, unmeasured worker attributes correlated with the percent female (e.g., cumulative work experience), and difference in women's and men's occupational labor supply preferences. Because the sex composition of an occupation is highly correlated with occupational measures of caring (see Table 4), an important question to address is how addition of a sex composition variable (the %Female) to wage regressions affect the coefficients on the caring variables (and vice versa). England et al. (2002) included %Female in their preferred specification. Given that it is negatively correlated with wages and positively correlated with caring work, they found that exclusion of %Female (as in our analysis) led to somewhat larger caring penalty estimates.

In order to save space, we summarize but do not report our results including %Female (they are available on request). Perhaps surprisingly, neither the cross-section nor longitudinal results are highly sensitive to inclusion of %Female in the wage equations. As expected, the wage level and wage change values of the caring coefficients become less negative after adding %Female, but the differences are modest and do not greatly change the interpretation of results. For example, in Table 5, the coefficients on the caring index for women and men (column 1) are −0.043 and −0.089; these decline (in absolute value) to −0.034 and −0.065 when %Female is added.

We also examine the reverse question. Do the estimated effects of % Female simply reflect wage differences associated with caring job attributes? Here we find that adding the caring attributes to wage level equations that include %Female does substantively decrease (in absolute value) the %Female coefficients, typically by about a third. As in Macpherson and Hirsch (1995), we find that %Female has a larger negative wage relationship for men than for women in wage level analysis, but that the magnitudes of the %Female coefficients drop sharply in longitudinal wage equations, being relatively small and similar for women and men. The longitudinal results indicate that a substantive portion of the wage effects associated with %Female stems for unobserved worker differences correlated with gender (e.g., cumulative work hours and experience). In a recent study, Ludsteck (2014) draws a similar conclusion using German administrative data that contain establishment as well as worker fixed effects.

In the larger academic and public discussion of gender wage differences, caring wage penalties are often proffered as an important (if not fundamental) determinant of the gender wage gap. Given the larger differences in levels of caring among jobs for women and men, such a focus would be appropriate if caring wage penalties were substantial (even more so if penalties were larger for women than men). That said, evidence in our paper indicates rather limited evidence for substantial market-wide caring penalties. Substantive estimates are found using wage level but not wage change analysis, yet the latter provides what are arguably more reliable results. The caring penalties that we do find are typically larger for men than for women.

How much might caring penalties narrow the gender wage gap? The difference in female and male means of the O*NET caring index is 0.56 (see Table 2). For sake of argument, assume that there exists a −0.02 wage penalty for a one standard deviation in caring. This roughly corresponds to the longitudinal estimates shown in Table 7 (column 3) showing caring coefficients of −0.014 for women and −0.022 for men. Multiplying 0.02 times the 0.56 caring difference indicates that the 0.190 gender wage gap in our sample (see Table 2)would be 0.011 log points lower absent the caring penalty, all else the same. Such a narrowing of the gender gap is not large. If one believes caring wage penalties are much larger, say, −0.05 rather than −0.02, the reduction in the gender gap would be less than 0.03 log points (0.05 times 0.56 = 0.028). Even if caring penalties were this large, policy implications are far from clear. What policies and shifts in attitudes might move us toward an alternative world in which we could eliminate caring wage differentials and/or differences in caring levels between women's and men's jobs? Are there not more feasible strategies through which gender equity can be enhanced and wage gaps narrowed?

Conclusion

Economists and sociologists have proposed plausible theories for why there may exist wage penalties for work involving helping and caring for others, jobs often performed by women. Previous evidence is mixed, but some studies have suggested wage penalties for women and men in caring jobs, on the order of 5%. Taken as a whole, the empirical evidence for caring wage penalties has been limited, varied, and not always convincing. Our expectation was that the use of large household samples of workers matched to multiple and varied measures of occupational “caring” would be likely to produce clear-cut and more compelling evidence of wage differentials associated with caring work.

Rather than designating occupations as caring or not, we use continuous measures from O*NET of “assisting & caring for others” and “concern for others,” plus four measures of “developing/teaching,” in order to construct alternative measures of caring in the labor market. Our principal results focus on a caring factor index that loads the “assisting& caring” and “concern” for others variables. We also construct broad-based measures of job skills/tasks and physical working conditions. All these job descriptors are based on evaluations from job analysts and incumbents for detailed occupations across the U.S. workplace. We match these job descriptors to multiple years of the Current Population Survey (CPS) earnings files, which enables us to perform both standard cross-section wage level and longitudinal wage change analyses, the latter based on large panels with two observations per worker (one year apart). Because longitudinal analysis identifies the effects of caring based on wage changes among workers moving into and out of occupations involving different levels of caring, it accounts for otherwise unobserved worker heterogeneity correlated with wages.

Although we do find evidence of wage penalties for caring work, the magnitudes of the penalties are modest or small. Using wage level analysis, we obtain estimates of caring wage penalties of roughly 4% for women and 8% for men resulting from a one standard deviation change in the caring index (these estimates would be zero for women and about 5% for men absent inclusion of our occupational skill index). For both women and men, the caring penalty is more strongly associated with a measure of "concern for others" than with the measure "assisting & caring for others." The former is intended to reflect needed worker attributes, whereas the latter reflects general work activities in the occupation. Our preferred estimates come from our longitudinal analysis (as in England et al., 2002), which accounts for worker heterogeneity. These estimates suggest caring penalties of 1.4% for women and 2% for men from a one standard deviation change in caring, with the "concern" and "assisting & caring" for others measures contributing similarly.

Jobs in the public sector involve substantially higher levels of caring for both women and men. When we provide separate wage level analyses for the public and private sectors, we find stronger evidence for caring penalties in the public than in the private sector for both women and men. That said, fewer than 1-in-6 wage and salary employees work in the public sector, although the rate is larger among women than men. We were unable to provide longitudinal analysis for the public sector given that occupation changes within the public sector are infrequent and would be poorly measured. Our expectation is that a substantive portion of the caring penalty estimates found in the public sector would reflect worker heterogeneity, just as we found for the economy-wide sample.

Also examined is how caring wage differentials vary across the earnings distribution. We do so using quantile regression and estimating OLS wage level and change equations across the quintiles of the predicted wage distribution. Although there are some differences in results between women and men based on the approach, the principal finding from the analysis is that caring wage effects are reasonably constant across the distribution.

There is a widespread belief among social scientists and the public that women are disproportionately employed in caring jobs and are penalized for doing so, receiving lower wages than they or similarly skilled workers would receive in jobs not requiring care. Our study confirms that the levels of “assisting & caring” and “concern” for others in the workplace are substantially higher for women than men. Estimates of caring wage penalties are sensitive to methods and specification, an expected result given that caring jobs include some of the most highly skilled and least skilled jobs in the economy. Although we find clear-cut evidence for lower wages in caring jobs using wage level analysis, the magnitude of these differences is sharply reduced using longitudinal analysis that accounts for worker heterogeneity. Our preferred estimates suggest caring wage penalties of about 2% for a one standard deviation increase in our caring index. Penalties are typically larger for men than for women, but fewer men work in high-caring jobs. We find little evidence for systematic differences in caring penalties across the wage distribution. The magnitude of estimated caring wage penalties is small as compared to other wage gaps in the labor market (e.g., union, industry, employer size, and city size wage differentials). Even were caring penalties erased from the labor market, there would be minimal closing of the gender wage gap.

II ARE STATES WINNING THE FIGHT? EVIDENCE ON THE IMPACT OF STATE LAWS ON BULLYING IN SCHOOLS

Introduction

Bullying is regarded as aggressive behavior intended to inflict harm on or control over another and is characterized by repetition and an imbalance of power (Olweus, 1997; 2003). Bullying takes many forms: physical (hitting, shoving, hand gestures, spitting on, throwing object at, taking things), verbal (name calling, taunting, threats), social/relational (rumor spreading, isolating from peers, embarrassing in public, purposeful exclusion), and electronic/cyber (using computers or cell phones to convey harmful words or images) (Olweus, 2003).

While there is much anecdotal evidence suggesting bullying has been around for decades (if not centuries), data measuring such historical accounts is limited.[31] Nevertheless, bullying has been recognized as a pervasive and persistent problem. Indeed, a number of large-scale health, crime, and education surveys added questions on students' experiences with bullying (e.g., HBSC since 1993; SCS/NCVS since 1999; SSOCS since 2000; YRBSS since 2009).[32] Table 14 provides a comparison of the prevalence of being bullied at school for the U.S. and international cross-sections in 2002 and 2010 based on estimates from the World Health

---

[31] Most adults can recall either being the bully or being bullied during their youth.

[32] The Health Behavior in School-Aged Children (HBSC) surveys students in grades 6-10 and is sponsored by the World Health Organization and administered in about 45 countries. The School Survey on Crime and Safety (SSOCS) surveys principals in K-12 schools and is sponsored by the National Center for Education Statistics. The School Crime Supplement to the National Crime Victimization Survey (SCS/NCVS) surveys students in grades 6-12 and is jointly sponsored by the National Center for Education Statistics and the Bureau of Justice Statistics. The Youth Risk Behavior Surveillance System (YRBSS) surveys students in grades 9-12 and is sponsored by the Center for Disease Control. SSOCS, SCS/NCVS, and YRBSS are administered strictly in the US and are discussed in more detail in the data section of this paper.

Table 14. Prevalence of Being Bullied at School, HBSC

	USA			International cross-section		
	2001/02					
	11 year-olds	13 year-olds	15 year-olds	11 year-olds	13 year-olds	15 year-olds
Bullied at school	35% girls 35% boys	36% girls 40% boys	26% girls 31% boys	35% girls 40% boys	34% girls 38% boys	25% girls 29% boys
Repeatedly bullied at school	12% girls 14% boys	12% girls 18% boys	7% girls 13% boys	13% girls 16% boys	12% girls 15% boys	8% girls 11% boys
	2009/10					
Repeatedly Bullied at school	13% girls 15% boys	12% girls 13% boys	7% girls 6% boys	12% girls 15% boys	11% girls 13% boys	7% girls 10% boys

Source: Author created using data reported in corresponding years of the WHO's International Reports, see http://www.hbsc.org/publications/international/

International cross-section is comprised of 35 countries in survey year 2001/02 and 38 countries in survey year 2009/10, both counts include the U.S. 'Bullied at school' refers to a student self-report of being bullied at school at least once in the previous couple of months. 'Repeatedly bullied at school' refers to a student self-report of being bullied at school at least two or three times in the previous couple months.

Organization's Health Behavior in School-aged Children (HBSC) International Reports.[33] Clearly, many students around the world report being bullied at school—a problem that does not appear to have subsided in the last decade. A similar pattern of prevalence and persistence in being bullied exists among adolescents in the U.S. Craig et al. (2009) point out there is marked variation in prevalence rates across the participating countries in the HBSC (not shown in Table 14) suggesting important differences in cultural norms and/or national policies.[34] Craig et al. (2009) point out there is marked variation in prevalence rates across the participating countries in the HBSC (not shown in Table 1) suggesting important differences in cultural norms and/or

[33] First, bullying is defined in the HBSC survey: "a student is being bullied when another student, or a group of students, say or do nasty and unpleasant things to him or her. It is also bullying when a student is teased repeatedly in a way he or she does not like or when he or she is deliberately left out of things. But it is not bullying when two students of about the same strength or power argue or fight. It is also not bullying when a student is teased in a friendly and playful way." Then adolescents were asked "how often they had been bullied at school in the past couple of months". Response categories ranged from "I was not bullied at school in the past couple of months" to "several times a week", see http://www.hbsc.org/publications/international/. The HBSC 2009/10 survey did not ask students to report whether they were bullied at school one or more times in the previous couple months.

[34] Craig et al. (2009) utilize HBSC 2005/06 data, but other survey years have similar patterns of variance as well, see http://www.hbsc.org/publications/international/.

national policies.[35] The authors also note that countries with a low prevalence of bullying involvement (mostly Scandinavian) have long-standing national programs that address bullying, but countries with a high prevalence (mostly eastern European) do not (Craig et al., 2009).

Estimates derived from national surveys illustrate that the persistently high prevalence rate of bullying is far more problematic than other behavioral issues facing U.S. schools. According to the School Crime Supplement to the National Crime Victimization Survey, roughly 28% of students in grades 6 through 12 reported being bullied at school in 2005, 2007, 2009, and 2011. Yet only about 5% of these students reported being in a physical fight at school and around 1% reported being absent because they felt unsafe.[36] Based on these estimates, more than seven million middle and high school students were bullied at school in 2011 alone.[37]

Bullying used to be "confined to the playground", but bullies' inappropriate use of technology literally follows victims into their homes and classrooms. In 2011, over 2 million middle and high school students in the U.S. (9% of students in grades 6-12) were electronically bullied in the last year, and 80% of these students were also bullied at school.[38] If schools provide little or no sense of sanctuary for victims of bullying, these students may skip class or be absent (Kochenderfer & Ladd, 1996; Greene, 2003; Glew et al., 2005); transfer, choose homeschooling, or drop out of school; or get into physical fights or resort to carry weapons in order to avoid further torment (Nansel et al., 2003). Consequently, measuring bullying would

---

[35] Craig et al. (2009) utilize HBSC 2005/06 data, but other survey years have similar patterns of variance as well, see http://www.hbsc.org/publications/international/.

[36] Estimates computed from SCS/NCVS 2005-2011 using sampling weights; see Figure 2 in text.

[37] Data on total enrollment in grades 6-12 taken from NCES ELSi for academic year 2010-2011.

[38] These figures were computed from SCS/NCVS data using sampling weights, see section on data for further discussion.

be a formidable task even if data were plentiful, accurately reported, and there was consensus in the population as to what constitutes bullying.

Federal departments have coordinated their efforts in order to disseminate information on bullying to the public (US GAO, 2012); an information campaign Stop Bullying Now! was launched in 2004 and a federally maintained website www.stopbullying.gov was introduced in 2011 during a White House conference on bullying (US GAO, 2012).[39] Additionally, the U.S. Department of Education and the U.S. Government Accountability Office released reports on the prevalence of bullying and states' legal efforts to combat it (Stuart-Casell et al., 2011; US GAO, 2012). Awareness among Congressional policymakers of the prevalence of bullying coupled with interest in providing universal and uniform coverage for students has led to several bills being introduced although no federal bullying legislation has been passed.[40]

In 1999, Georgia became the first state to put any type of bullying legislation into effect. Since then many states enacted bullying laws, most going into effect in the mid- to late-2000's. As of April 2012, only one state (Montana) had no school bullying legislation in place (US GAO 2012). In addition to variation across states in the timing of adoption, there is also variation across states in the type of bullying legislation. While some states' legislation mandates school districts to take measures against bullying, it is less clear whether the laws ensure that schools comply with state mandates particularly if a compliance deadline is not specified.[41] So I also

[39] The three federal departments are the Department of Education (ED), Department of Health and Human Services (HHS), and Department of Justice (DOJ). Other responses included creating the Federal Partners in Bullying Prevention Steering Committee in 2009 to promote collaboration on bullying across the Federal Government.
[40] See US GAO (2012) footnote 5 for some examples of bills related to school bullying.
[41] In 2012, New York put a bullying law into effect that mandated schools report incidents to the state's Education Department. Recent media coverage exposed that 51.4% of public schools in New York State reported zero incidents of harassment and discrimination (and more than 60% reported zero instances of cyberbullying) during the 2013-2014 school year to the state's Education Department (Spewak, 2015).

examine whether the type of legislation matters by classifying states' laws into a broad class referred to as "any" and two sub-classes referred to as "mandate" and "mandate-deadline".

Determining whether, and to what extent, states' laws reduce bullying in schools is important for several economic reasons. First, understanding the impact of state legislation on in-school bullying can inform proposed federal legislation that aims to unify state laws. Second, bullying can be financially problematic for schools. Bullying increases operating costs if administrative time is spent responding to incidents (i.e., investigation, disciplinary action, documentation); if vandalism to school property occurs (destroying text books, damaging lockers, etc.); or if indicted bullies need alternative educational placements. Schools can suffer losses in revenue because low attendance due to bullying (i.e., increased number of truancies, health-related absences, and dropping out for victims; increased numbers of suspensions and expulsions for indicted bullies) reduces Average Daily Attendance rate reimbursements from the state. Third, an economics literature suggests bullying might have serious negative labor market consequences by disrupting both cognitive and non-cognitive skill development (e.g., Brown and Taylor, 2008; Powdthavee, 2012; Ammermueller, 2012; Eriksen et al., 2013).[42] Lastly, to the extent that schools are incubators for the workforce, it is believed that bullying among youths may carry over into adulthood and have longer-run impacts within households, labor markets, and the larger economy.

This paper aims to enhance the evidence base on legislative policies implemented to reduce bullying among youths by empirically answering the question: Do states' bullying laws

[42] If being bullied as a youth affects one's mental health status (e.g., attitude and self-esteem) and if certain psychological traits are linked to lower educational attainment and earnings, then one's labor market potential could be diminished by being bullied. Waddell (2006) finds that high school students possessing poor attitudes and low self-esteem have lower educational attainment and thus a higher probability of unemployment and lower earnings as adults.

abate bullying in schools? To do so, I perform both student-level and school-level policy evaluations. First, I pool biennial cross-sections of student-level data from a large, nationally representative survey of high school students spanning years 1993-2011. These data are used to estimate the effect of states' bullying laws on the probability that a student will report having been bullied or having experienced bullying-related behaviors at school. A difference-in-differences (DD) approach is used to exploit variation across states in the timing of adoption and type of bullying legislation. Focusing on the impact of having any type of bullying law in effect, I estimate both two-period and dynamic policy effects. The former captures the average effect of bullying laws in a pre- and post-policy setting, while the latter captures the average policy effects over time (i.e., separate estimates for years before and after the effective year) and allows me to investigate concerns about policy endogeneity and persistence effects. I also compare estimates across different types of legislation (any, mandate, mandate-deadline) using control states (i.e., states' never having bullying legislation in effect during the sample) as the comparison group. Then in a similar analysis, I estimate the effects of states' bullying legislation using pooled biennial cross-sections of school-level data from a nationally representative survey of principals. These data span the years 2004-2010 and provide bullying outcomes plus transfer rates for the K-12 population. Here I compare estimates of the impact of states' bullying laws by type and by grade level group (elementary, middle, high school). The latter comparisons enable me to assess the external validity of the impact estimates from the student-level analysis, which are based on data from high school—a population markedly less afflicted by bullying in schools.

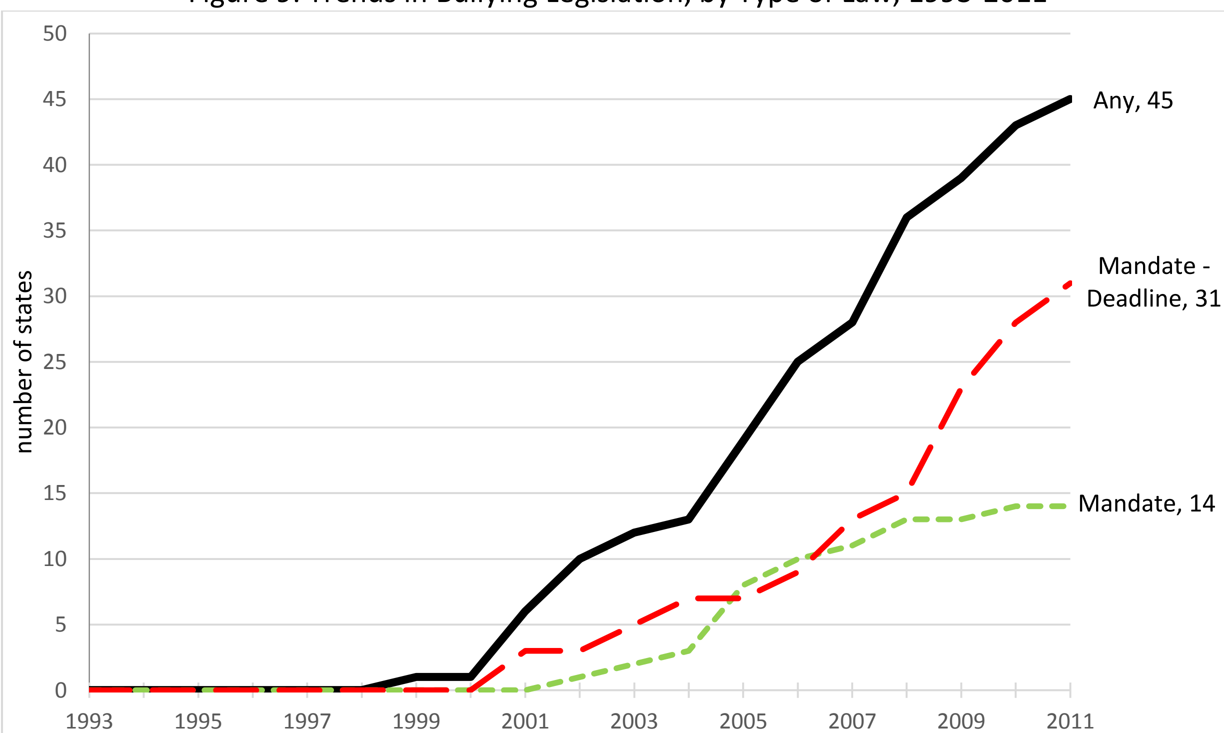
States' Legislative Efforts to Reduce Bullying in Schools

Terminology, definitions, governance over schools' bullying prevention, and school accountability for reporting and responding to incidents of bullying are just a few of the components of bullying legislation that vary widely across states (Stuart-Cassell et al., 2011; US GAO, 2012). Moreover, some states may explicitly direct schools to implement bullying prevention programs and/or an administrative protocol for responding to and reporting incidents of bullying using a state model policy, while other states' bullying laws mandate little to no action at the school level (Stuart-Cassell et al., 2011; US GAO, 2012). In fact, this variation—the degree to which a state's legislation filters down to affect school districts' actions and/or responses to bullying—is likely to have the greatest influence on the effectiveness of states' laws at reducing bullying in schools. In order to examine whether the impact differs by the type of bullying legislation, states' laws are classified in the following way.

Analysis of the bullying legislation should begin with a basic class that indicates whether a state has effected any type of bullying law whatsoever regardless of its coverage; I refer to this type of bullying law as "any". But a state-level bullying law that prohibits bullying is more likely to affect students' experiences in schools if the bullying law mandates school districts to develop and implement school policies on bullying; this criterion constitutes the "mandate" class of bullying laws. If no date is specified by which districts must comply with a state bullying law, however, many schools may fail to do so because of a lack of additional funding from or monitoring by the state. The "mandate-deadline" bullying law criterion refers to a class of bullying laws that requires school districts to comply with a state's mandate for a local policy by a specified deadline. One would expect the "mandate-deadline" type of legislation to have an impact on school bullying that is stronger in magnitude than any impact of "mandate" only type. Whereas the magnitude of the impact of having "any" type of legislation is likely somewhere in between since this class is comprised of both types and a few bullying laws that cannot be classified as either.[43]

Figure 9 depicts trends in states' effective year of bullying legislation by type providing some insight into the amount of variation in timing and type of law. Most of the variation occurs in the mid- to late-2000's, overlapping well with the data used in the evaluation.

Figure 9. Trends in Bullying Legislation, by Type of Law, 1993-2011



Source: Author created based on information from US DOE report and states' legislative websites.
The graph depicts the year in which a state's bullying legislation was effective as opposed to adopted or enacted. See section 2 in text for further discussion on classifying states' bullying laws as "any", "mandate", or "mandate-deadline."

[43] Only a few states have such legislation, which typically prohibits bullying in schools and may provide a definition of bullying behavior.

Economic Framework for Analyzing State Bullying Legislation

State bullying legislation typically does not recognize bullying as a crime per se like laws governing incidents of assault or statutory rape. Having a state bullying law in effect, however, may affirm bullying is a social injustice. In this sense, one could take an economic approach similar to Becker's Economics of Crime (1968) to understand the mechanisms through which state legislation may impact in-school bullying. That is, how do state bullying laws affect the incentives (benefits and costs) facing bullies and their victims.

If a bullying law requires no action by school districts and merely prohibits bullying and possibly provides a definition of bullying behavior, then bullying another student is essentially free. Victims may also recognize that the laws pose no credible threat to bullies. Thus, it is likely that such laws may have no effect on either students self-reports or principals accounts of student bullying in schools.

Alternatively, a bullying law could require school districts to take action against bullying including (i) explicitly outlining bullying behavior and school policies on bullying in a student handbook/code of conduct, (ii) implementing bullying prevention programs, and/or (iii) adopting and enforcing strict disciplinary actions against bullying in school. Laws that heighten awareness of bullying behavior among students may empower victims to speak out by lowering the social costs to victims for reporting incidents of bullying to administration. If so, the laws may increase both students self-reports and principals accounts of in-school bullying. Arguably, a heightened awareness about bullying could also reduce incidents of bullying if it increases the social costs to bullies for bullying other students. Laws mandating that districts enforce strict disciplinary action may increase the expected costs for bullies by posing credible threats for bullying other students.[44] These laws likely reduce incidents of student bullying and thereby students self-reports of being bullied at school may decrease; however, the effect of these laws on principals accounts of student bullying is unclear, depending on whether the deterrent effect on bullies dominates any empowerment effect on victims.

Data

There are two large, representative cross-sectional surveys measuring the bullying-related experiences of American youths in schools. The national Youth Risk Behavior Surveillance System (YRBSS) and the School Crime Supplement to the National Crime Victimization Survey (SCS/NCVS) are biennially administered and span years before and after the effective dates of states' bullying legislation.[45] Only the YRBSS has state-identifiers for multiple survey years in a format available to researchers.

Bullies or victims of bullying might transfer schools, opt for home-schooling, or drop-out of school completely (Le et al., 2005), however, the YRBSS does not provide such measures. Moreover, determining whether states' laws reduce bullying in a school as a whole is equally informative as determining whether states' laws reduce bullying for a particular student. The School Survey on Crime and Safety (SSOCS) provides school-level measures of bullying and transfer rates. These three data sets are discussed below.

---

[44] The following two small case studies provide examples. Raynor & Wylie (2012) find students who reported that their school had a bullying policy and took action against bullying were more likely to speak out against bullying. Frisén and co-authors (2012) find "the most frequent answer to the question about what made bullying stop was that the bullying ended with the intervention of school personnel." Moreover, a meta-analysis of the impacts of anti-bullying programs in schools performed by Ttofi & Farrington (2011) also finds school disciplinary methods play an important role in reducing student bullying.

[45] Other data sets are not well suited to measure the effect of state bullying laws. The Health Behavior in School-aged Children (HBSC) is administered only every four years and access to data for publication is limited to HBSC network members only. The National Survey of Children's Exposure to Violence (NatSCEV) was administered only in 2008. Waves I (1994-1995) and II (1996) of the National Longitudinal Study of Adult Health (Add Health) questions adolescents in middle and high school grades, but these survey years occurred before any states had bullying legislation in effect (i.e., prior to 1999).

Youth Risk Behavior Surveillance System (YRBSS)

The national YRBSS is a school-based survey conducted in the spring of odd years by the Center for Disease Control (CDC) since 1991. It is administered to a representative sample of about 15,000 students in grades 9 through 12 in public and private schools across the U.S. in order to monitor “health-risk behaviors that contribute to the leading causes of death and disability among youth and adults” (CDC, 2013). Note that the same states do not participate in the YRBSS each survey year and the schools within each state also vary across survey years. Appendix Table B1 lists the states surveyed in the national YRBSS for the years 1993-2011.

I utilize relevant (dichotomous and categorical) variables in the Unintentional Injuries and Violence section of the national YRBSS survey.[46] There are two dichotomous variables in the YRBSS that explicitly measure whether a student was bullied. ‘Bullied on school property during the 12 months before the survey’ was included in the 2009 and 2011 surveys and ‘ever been electronically bullied (including through e-mail, chat rooms, instant messaging, Web sites, or texting)’ was included in 2011 survey. Ideally, one would like to have several survey years where bullying and electronic bullying questions where included. Since difference-in-differences identification is based on only two states (one switching from no law to a “mandate” law; one switching from no law to a “mandate-deadline” law), estimates of the impact of bullying laws on ‘bullied at school’ are unreliable and as such are omitted from the tables though discussed in the text. Having only a single cross-section precludes estimating the impact of bullying laws on ‘ever been electronically bullied.’

[46] I use protected-use versions of the national YRBSS data with state-identifiers, which were acquired directly from the CDC following procedures outlined on their website.

Consequently, I use other relevant (dichotomous and categorical) outcome variables measuring whether a student experienced bullying-related behaviors at school. Students who were bullied at school in years when the YRBSS did not ask about being bullied per se (i.e., prior to 2009) likely partook in avoidant or violent behaviors at school, which may have been correlated with being bullied. These other bullying-related variables consist of 'did not go to school because they felt unsafe at school in past 30 days'; 'in a physical fight on school property one or more times in past 12 months'; 'threatened or injured with a weapon on school property one or more times in past 12 months'; and 'carried a weapon on school property in past 30 days'.[47] Students can respond to one of five categories measuring the frequency of 'absent because felt unsafe' or 'carried a weapon at school' in the past 30 days with the options ranging from 0 days to 6 or more days; they can respond to one of eight categories measuring the frequency of 'in a fight at school' and 'threatened at school' in the past 12 months with the options ranging from 0 times to 12 or more times.[48] Appendix Table B2 shows the survey years in which bullying-related questions were asked, all four were included in odd years 1993-2011. An advantage of using the bullying-related variables in the YRBSS is their dichotomous measures allow one to estimate the impact of bullying laws at the extensive margin while their categorical (frequency) measures allow one to estimate the impact at the intensive margin. A disadvantage of using these bullying-related variables, however, is that they measure both delinquency and bullying (i.e., bullying behaviors deemed as bullying others and being bullied).

[47] 'Having had property stolen or damaged at school in the past 12 months' is another bullying-relevant question, however, this question is not asked consistently throughout the sample period so it is excluded from the analysis.
[48] The YRBSS provides dichotomous measures based on these categorical responses in later survey years and one can create dummy variables in earlier survey years that do not provide dichotomous measures.

Additionally, I use the frequency measures of the bullying-related outcomes to create dichotomous measures of whether a student repeatedly partook in avoidant/violent behaviors. That is, I construct four additional binary outcome variables each equal to one if a student reported experiencing the bullying-related behavior two or more days/times in the corresponding period: 'repeatedly absent because felt unsafe at school', 'repeatedly in a physical fight on school property', 'repeatedly threatened or injured with a weapon on school property', and 'repeatedly carried a weapon on school property'.[49]

I expect these bullying-related measures to be less responsive to legislation than an explicit and comprehensive measure of in-school bullying because they also capture other student behaviors such as delinquency. That is, estimates of the effect of states' legislation on students' bullying-related experiences will likely understate the impact of states' laws on in-school bullying.

School Climate Survey/ National Crime Victimization Survey (SCS/NCVS)

The SCS/NCVS is a national survey of about 6,500 students ages 12-18 in public and private schools. It is conducted in 1989, 1995, 1999, 2001, 2003, 2005, 2007, and 2009 and is jointly sponsored by the National Center for Education Statistics and the Bureau of Justice Statistics.[50] The SCS is a supplement to the NCVS and collects information about victimization, crime, and safety at school.

---

[49] While repetition could be the same behavior or some combination of bullying behaviors, I also construct a binary bullying-related index that equals one if a student responds affirmatively to two or more of any combination of the four dichotomous bullying-related measures: absent because felt unsafe, in a fight at school, threatened at school, and carried a weapon at school. I do show these results, however, since 'absent' and 'weapon' are measured in days while 'fight' and 'threatened' are measured in times and their reference periods differ as well. Thus, it is unclear how one would interpret any policy effects for such a bullying-related outcome measure.

[50] I use public-use versions of SCS/NCVS data without geographic-identifiers, which were acquired from the ICPSR following procedures outlined on their website (citations in Reference section). SCS has state-identifiers available

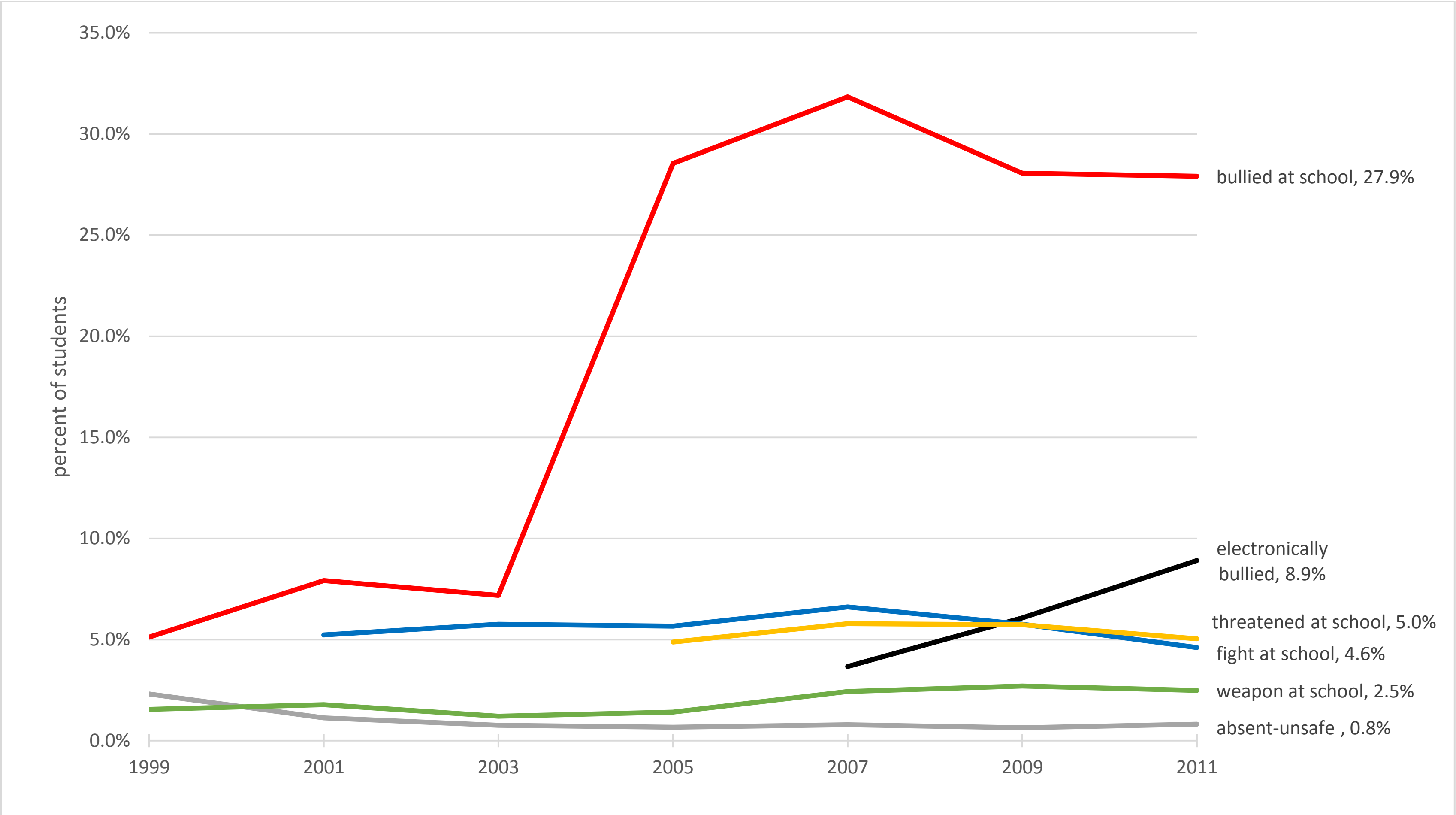
The SCS/NCVS asks whether a student was bullied at school (since 1999), how often the bullying occurred (since 2001), and what specific form of bullying (verbal, relational, physical, etc.) occurred (since 2005). But, unlike the YRBSS, the exact wording of the 'bullied at school' question changes twice. Appendix Table B3 shows how the language used in the 'bullied at school' question evolves over time. In 2005, the 'bullied at school' definition changes from a rather simple definition describing coercion and verbal bullying acts into a comprehensive definition incorporating seven types of bullying. In 2007, the reference time period for the SCS was revised to include incidents in "this school year" instead of in "the last six months" because "respondents revealed they were not being strict in their interpretation of the 6-month reference period and were responding based on their experiences during the entire school year" (DeVoe & Bauer, 2010).[51] Figure 10 shows the national trends in the prevalence of bullying experiences of youths aged 12 to 18 as reported by students surveyed in the SCS/NCVS. One might expect prevalence rates for being bullied at school to be higher as the definition of bullying becomes more inclusive or covers a longer period of time. Indeed, there is a sharp increase in the percent of students in grades 6-12 being bullied at school from 7% in 2003 to 29% in 2005 (following the expanded definition of bullying) followed by another slight increase to 32% in 2007. In fact, most bullying-related experiences had a slight increase in 2007, however, the 2007 jump does not appear to persist in 2009 and 2011. This could be explained by the rotational group design of the NCVS where roughly 25% of SCS/NCVS respondents in

---

in electronic format only for survey years 2009 and 2011 through restricted-use terms at a Census Bureau Research Data Center.

[51] The change in wording in the 2007 SCS survey questions was prompted by cognitive laboratory evaluations conducted by the Census Bureau (DeVoe & Bauer 2010).

Figure 10. National Trends in Bullying-related Experiences among Middle and High School Students, SCS/NCVS 1999-2011



Means are calculated using sample weights provided in the SCS/NCVS for years 1999-2011. 'In a fight at school' was included in survey years 2001-2011; 'electronically bullied' was included in survey years 2007-2011; 'threatened at school' is a particular form of bullying included in survey years 2005-2011. 'Bullied at school' became more comprehensive in 2003 and changed in 2007 to include experiences in past academic year (instead of 6 months), see text and Appendix Table B3 for further details.

2007 were also surveyed in 2005 but they would not have been surveyed in 2009 or beyond.[52] Nonetheless, I use caution in comparing the descriptive evidence from earlier survey years to estimates from later SCS survey years.

In addition, the SCS/NCVS includes demographics controls (age, grade, sex) and ratings of school climate (5) on a four-category scale, dichotomous measures of school safety (6 for years 1995, 1999; 9 for years 2001-2011), and categorical measures of household income.

School Survey on Crime and Safety (SSOCS)

The School Survey on Crime and Safety (SSOCS) is a nationally representative cross-sectional survey administered to about 3,500 public school principals in schools consisting of all grade levels (i.e., elementary, middle, high, and combined). It provides estimates of school-wide programs and policies related to crime prevention and reported incidents of crime, discipline, and disorder. The survey is conducted in late spring and I pool data from academic years 2003/04, 2005/06, 2007/08, and 2009/10.[53] Every state participates in the SSOCS in all survey years, but the schools within each state vary across survey years.

In the SSOCS the frequency of student bullying at a school is a qualitative measure from the set of 5 categories: happens daily, happens at least once a week, happens at least once a month, happens on occasion, and never happens. Respondents are predominantly principals and are asked to respond based on “the best of [their] knowledge.” So it is unclear whether a respondent reflects on actual recorded or reported incidents of student bullying at their school.

[52] The NCVS is administered every six months for a total of seven times per respondent 12 years of age or older in a surveyed household (initial plus six follow-ups), while the SCS is a biennial supplement to the NCVS. Thus, a youth aged 12-18 participating in the NCVS could participate in SCS in two consecutive biennial survey years at most.
[53] I use restricted-use versions of SSOCS data with state-identifiers, which were acquired from the NCES following procedures outlined on their website; 1999/2000 survey year did not included state FIPS codes so it was excluded.

And one may suspect that this holistic measure might not align with (and underestimate) unreported incidents.

Other bullying-related school-level measures are also included in the SSOCS. Average daily attendance (ADA) rate and the number of students transferred to the school, the number of students transferred from the school, the number of students involved in physical attacks/fights, and the number of students involved in using or possessing a weapon. I divide each of the latter four variables by total student enrollment to create four bullying-related outcome variables: transfer-IN rate, transfer-OUT rate, fight rate, and weapon rate. ADA rate is a percent expressed in decimal form, but the other four rates refer to their corresponding count variable as a share of total student enrollment, each is a non-negative continuous outcome variable.[54]

In addition, the SSOCS includes measures of school/student demographics (percent of students who are male, belong to a specific race/ethnic group, or do not speak English as their first language), parents' economic resources (percent of students eligible for free/reduced price lunch), grade levels served, crime level of the area where the school is located, pupil-teacher ratio, and total student enrollment.[55]

---

[54] ADA rate has weighted sample mean 0.9376, standard deviation 0.0792, a minimum of 0.01 and a maximum of 1.00 where 85% of the weighted sample has an ADA rate between 0.90 and 0.98, inclusive.

[55] SSOCS includes frame variables for the number of white, black, Asian, American Indian, and Hispanic students. These frame variables are usually taken from the Common Core of Data (CCD) from two years prior to the SSOCS survey year and merged to the SSOCS data by a school-specific identifier. I divide these two-year lagged counts by current year total student enrollment in order to construct school-level measures of percent black, percent Hispanic, percent white, and percent other race. While not an ideal measure, racial/ethnic composition of a school may explain an important amount of variation in the outcome variables; and it is unlikely that variance in the measurement error in school racial/ethnic counts (difference between the true survey year count and the two-year lagged count) is systematically correlated with the policy variable or the outcome.

Additional Independent Variables

State-level covariates are collected from various sources to supplement student demographic and academic characteristics provided in the YRBSS and improve the policy evaluation by increasing the statistical power of the primary, or student-level, analysis (i.e., increasing the precision of the impact estimates). In the event that bullying might be systematically related to parents' economic resources, I control for median household income and unemployment rate of a state.[56] Since bullying might be affected by community climate, I control for violent crime rate and property crime rate per 100,000 population.[57] I also control for the size of state using ln(population). Some school quality indicators may also play a role in reducing the incidents of bullying: lower pupil-teacher ratios may mean more supervision or smaller schools, and higher total per pupil expenditures may be related to more funding for school safety measures or school programs aimed to improve school climate.[58] Note corresponding measures already exist in the school-level (i.e., SSOCS) data so I do not include state-level covariates in my secondary analysis.

Ideally, one would like to obtain two additional state-level controls that may be correlated with both the policy variable and the bullying outcome of interest. The first concerns compliance rates within a state. Even when state legislation prescribes that a state-level education governing body monitor that schools comply with the bullying law, there is still the possibility that a school may not act in accordance with the bullying protocol they submitted to

[56] Median household income was gathered from the US Census Bureau website and unemployment rate from the Bureau of Labor Statistics website by state and by sample year.
[57] Crimes rates and population by state and by sample year were collected from the US Department of Justice, FBI, Uniform Crime Report. 2010 figures were the most recent year available, so we used 2010 figures for sample year 2011.
[58] Pupil-teacher ratio, total per pupil expenditure, and the fraction of students in schools in grades 9-12 were collected from NCES Common Core of Data "Build a Table" website.

the state authority. But no data set is available that explicitly tracks school compliance rates with state bullying laws.

Second, not all states' bullying laws address bullying in schools in the same way or to the same degree, so schools may undertake preventive measures on their own that directly or indirectly affect bullying. Other initiatives intended to improve school climate (e.g., bullying awareness training for faculty and staff, cyber-bullying programs for the community) may also influence bullying-related behaviors in schools. Ideally one would like to control for whether a student surveyed attends a school that participates (or had previously participated) in a program to improve school climate. While the SSOCS includes principals' accounts of having such programs or policies measures at their school (typically dichotomous variables), no student-level survey with state-identifiers contains such measures.[59] Furthermore, there is no survey that consistently collects data on the percent of all schools in a given state and year that currently or previously participated in a bullying prevention program.[60]

Methodology

Student-level

I employ a difference-in-differences (DD) approach to estimate the effect of bullying laws on a student's probability of reporting being bullied at school. Equation (1) gives the regression model that I estimate by OLS using pooled cross-sections of YRBSS data merged to data on states' bullying laws for odd years 1993-2011. Sampling weights provided in the YRBSS are used.

[59] SCS/NCVS contains students' accounts of school climate and some school policies and security measures.

[60] School Health Profiles (SHP), a national survey conducted by the Center for Disease Control (CDC), provides cross-sectional data for these measures and is publically available at state-level aggregation. While SHP is administered to school principals and lead health teachers every even year from 1996 to 2010, it only intermittently asks about bullying prevention programs in schools during the sample period.

$$y_{ist} = \alpha + \gamma policy_{st} + \beta_1 X_{ist} + \beta_2 W_{st} + \delta_1 state_s + \delta_2 year_t + \varepsilon_{ist} \quad (1)$$

y_{ist} denotes the binary student-level outcomes (bullied on school property; absent because felt unsafe; in a fight at school; threatened at school; carried weapon at school; and their "repeated" counterparts) of student *i* in state *s* in year *t*. Note that each dependent variable equals one if a student responds affirmatively to the particular question asked in the YRBSS survey. X_{ist} is a vector of student-level demographic covariates (sex, age, grade level, race, ethnicity), W_{st} is a vector of state-level covariates in state *s* and year *t* (unemployment rate, median household income, crime rates, school quality indicators, log population), and ε_{ist} is an idiosyncratic error term. Cluster-robust standard errors are estimated using clustering at the state level to address concerns of serial correlation (Bertrand et al., 2004; Cameron & Miller, 2015). $state_s$ is a set of state dummy variables (i.e., fixed effects) that control for unobserved time-invariant state-specific characteristics correlated with both the bullying outcome and having or not having a given type of bullying law in effect (e.g., political, cultural or religious attitudes towards the role of schools in preventing bullying such as community outreach or parental involvement). $year_t$ is a set of year fixed effects that control for unobserved time-varying characteristics common to all states and correlated with both the bullying outcome and having or not having a given type of bullying law in effect (e.g., national media coverage about student bullying in schools).

Replacing the binary student-level outcomes with their ordered categorical frequency measures, I estimate ordered probit regression models given by equation (2). Sampling weights provided in the YRBSS are used.

$$y_{j,ist} = \alpha + \gamma policy_{st} + \beta_1 X_{ist} + \beta_2 W_{st} + \delta_1 state_s + \delta_2 year_t + \varepsilon_{ist}\ ,\ \ j \in \{1, 2, \dots , J\} \quad (2)$$

Here $y_{j,ist}$ denotes the *j*th level of frequency for the bullying-related outcome (absent because felt unsafe; in a fight at school; threatened at school; carried weapon at school) of student i in state s in year t. Note that 'absent' and 'carried weapon' each have J = 5 possible frequency categories (0 days, 1 day, 2-3 days, 4-5 days, 6 or more days) while 'fight' and 'threatened' each have J = 8 possible frequency levels (0 times, 1 time, 2-3 times, 4-5 times, 6-7 times, 8-9 times, 10-11 times, 12 or more times). X_{ist}, W_{st}, state and year fixed effects, and the error term are the same as described above.[61]

Regardless of whether the outcome is a binary or categorical variable, the key covariate of interest is a dummy variable, $policy_{st}$, that indicates whether state *s* has a bully law that is in effective in year *t.* Thus, γ captures the two-period policy effect (any, mandate, mandate-deadline). The policy variable is constructed using information on states' laws taken from a report submitted to the U.S. Department of Education (Stuart-Cassell et al., 2011), then cross-validating and updating these dates using states' online legislation databases.[62]

I estimate the two-period policy effect for each type of legislation (any, mandate, mandate-deadline) in separate analyses but using the same comparison group—control states. Control states are states that never put any type of bullying legislation in effect during the YRBSS sample period (1993-2011) conditional on having participated in the YRBSS. Recall that

---

[61] The error term in Equation (2) is assumed to be Normally distributed, hence, the choice of a probit (and not logit) regression model.

[62] I compiled dates for each of the three bullying law classifications for each state using Appendix B in USDOE report (Stuart-Cassell et al., 2011) and cross-checking each state statute with the state's legislative database in order to obtain the effective date of the bullying law. This process allowed for an accurate classification of each statute as either any, mandate, or mandate-deadline bullying law and served to familiarize me with the variation in coverage.

every bullying law belongs to the “any” category, while “mandate” and “mandate-deadline” are mutually exclusive sub-categories of “any”. When estimating the impact of, say, “mandate-deadline” laws, only observations from control states and from “mandate-deadline” states are included in the sample.

To estimate the dynamic effects of states’ laws on bullying in schools, I replace the single policy variable with a set of policy variables such that the policy effect (whether any, mandate, or mandate-deadline) has a time series dimension. Letting 1-2 years before the law be the omitted reference group, this set of policy variables includes dummies for the year the law was effected (i.e., year 0), 1-2 years after the law, 3-4 years after, 5-6 years after, 7 or more years after, 3-4 years before the law, 5-6 years before, and 7 or more years before. Estimating the dynamic policy effects allows one to empirically investigate whether there is any evidence of policy endogeneity (significant, non-zero impacts before the law was effective) and whether there is any evidence of persistence (significant, non-zero impacts after the law was effective).

School-level

I also employ a difference-in-differences (DD) approach to estimate the effect of bullying laws on a principal’s report of the school-level outcome of interest. To do so, pooled cross-sectional SSOCS data is merged to data on states’ bullying laws for the sample period consisting of even years 2004-2010. I use OLS to estimate the regression model given by equation (3) for all continuous bullying-related outcomes (i.e., rates). Sampling weights provided in the SSOCS are used.

$$y_{cst} = \alpha + \gamma policy_{st} + \gamma_M(policy_{st} \times Middle_{cst}) + \gamma_E(policy_{st} \times Elementary_{cst}) + \gamma_C(policy_{st} \times Combined_{cst}) + \theta_M Middle_{cst} + \theta_E Elementary_{cst} + \theta_C Combined_{cst} + \beta_1 X_{cst} + \delta_1 state_s + \delta_2 year_t + \varepsilon_{cst} \qquad (3)$$

y_{cst} denotes the school-level outcome (ADA rate, transfer-IN rate, transfer-OUT rate, fight rate, and weapon usage/possession rate) of school *c* in state *s* in year *t,* all of which are continuous variables. X_{cst} is a vector of school-level covariates (% female, % black, % Hispanic, % other race, % free/reduced price lunch, and dummies for low crime level where school is located). There are three dummy variables for grade level group where high school is the omitted (or base) category: $Middle_{cst}$, $Elementary_{cst}$, and $Combined_{cst}$. $state_s$ and $year_t$ are state and year fixed-effects and ε_{cst} is an idiosyncratic error term as described for equation (1).

I use an ordered probit regression model to estimate the model given by equation (4) for the frequency of student bullying, an ordered categorical outcome variable. Sampling weights provided in the SSOCS are used.

$$y_{j,cst} = \alpha + \gamma policy_{st} + \gamma_M(policy_{st} \times Middle_{cst}) + \gamma_E(policy_{st} \times Elementary_{cst}) + \gamma_C(policy_{st} \times Combined_{cst}) + \theta_M Middle_{cst} + \theta_E Elementary_{cst} + \theta_C Combined_{cst} + \beta_1 X_{cst} + \delta_1 state_s + \delta_2 year_t + \varepsilon_{cst}\,, \quad j \in \{1, 2, 3, 4, 5\} \qquad (4)$$

Here $y_{j,cst}$ denotes the *j*th level of frequency for the school-level measure of student bullying for school *c* in state *s* in year *t,* where 1 denotes never happens, 2 is on occasion, 3 is at least once a month, 4 is at least once a week, and 5 is almost daily. $Middle_{cst}$, $Elementary_{cst}$,

$Combined_{cst}$, X_{cst}, W_{st}, state and year fixed effects, and the error term are analogous to those described above for equation (3).[63]

For both equations (3) and (4), cluster-robust standard errors are estimated using clustering at the state level to deal with issues of serial correlation (Bertrand et al., 2004; Cameron & Miller, 2015). Also note that state-level covariates by year are not included here because the SSOCS includes school-specific data related to socio-economic status and local crime levels that the YRBSS student-level data lacks and for which the state-level covariates serve as proxies.

There are four policy variables of interest: a dummy variable, $policy_{st}$, that indicates whether state *s* has a bully law that is in effective in year *t* and three policy interaction terms: $policy_{st} \times Middle_{cst}$, $policy_{st} \times Elementary_{cst}$, and $policy_{st} \times Combined_{cst}$. Thus, γ captures the two-period policy effect for high schools while γ_M, γ_E, and γ_C captures the two-period policy effects for middle, elementary, and combined grade schools, respectively. These estimates allow me to assess the external validity of estimates based on data from the YRBSS. It is interesting to do so since prevalence rates of bullying are markedly higher in middle school than in either high school or elementary school (see descriptive and empirical evidence below) and the laws may impact students differently across these grade level groups.[64]

I estimate the policy effects (any, mandate, mandate-deadline) separately using the same comparison group—control states. Controls states are states never having put bullying legislation into effect during the SSOCS 2004-2010 sample period. As in the student-level

---

[63] The error term in Equation (4) is assumed to be Normally distributed, hence, the choice of a probit (and not logit) regression model.

[64] To the best of my knowledge, there is no nationally representative student-level data set that measures bullying among students in grades pre-kindergarten/kindergarten through grade 6 (i.e., elementary school).

evaluation, when estimating the “mandate-deadline” policy effect only observations from states with “mandate-deadline” bullying laws and control group states are included in the analysis; observations from “mandate” states do not become part of the control group.[65]

Because the SSOCS sample period consists of even years from 2004 to 2010 and most of variation in the states’ laws occurs in the mid- to late-2000’s, estimating the dynamic policy effects (essentially using a set of leads and lags in the policy variable) is not reasonable. The short time span of the SSOCS sample period reduces the number of years before and after the effective year of the bullying law for many states, reducing the power to detect dynamic policy effects in the school-level evaluation of bullying laws.

Results

Descriptive Evidence at the Student-level

Between 20% and 28% of adolescents were bullied at school during the last 12 months in the U.S., consonant with estimates based on periodic international surveys on the prevalence of being bullied in schools.[66] Being bullied at school affects far more American adolescents on average than any other related experience including fighting at school (12% YRBSS; 5% SCS/NCVS) and absenteeism due to feeling unsafe in school (6% YRBSS; 1% SCS/NCVS).[67] Considering that such violent and avoidant measures likely capture both delinquent and bullying behaviors underscores that bullying per se has been a major problem in U.S. schools in recent years.

---

[65] Similarly, when estimating the “mandate” policy effect, observations from states with “mandate-deadline” bullying laws are excluded from the analysis—they do not become part of the control group.

[66] 20% is based on national YRBSS data from 2009 and 2011; 28% is based on SCS/NCVS data from 2005-2011. Appendix Figure B1 depicts national trends in bullying experiences for high school students based on data from the YRBSS; recall Figure 2 depicts national trends in bullying experiences for middle and high school based on data from the SCS/NCVS.

[67] Here I compare estimates based on data from 2011 survey year.

There are important differences in bullying experiences between middle and high school students. Based on SCS/NCVS data from survey years 2005-2011, the prevalence of being bullied at school is higher among middle school students (35%) than among high school students (25%).[68] In fact, mean prevalence rates for six out of seven types of bullying described in the survey are statistically higher for middle school students.[69] Verbal (made fun of, name-called, insulted), relational (rumors, persuade others to dislike), and physical (pushed, shoved, tripped, spit-on) were the most commonly reported forms of being bullied experienced among students in grades 6-12 at 17%, 17%, and 9%, respectively. Yet high school students are half as likely to experience physical bullying (6%) and property destruction (2.6%) at school as compared to middle school students (14% and 4.7%, respectively). Prevalence rates of cyberbullying are slightly higher for high school (7%) than middle school students (6%).[70] Such marked differences may lead one to conjecture that states' bullying legislation could have different impacts on subgroups of students (i.e., elementary, middle, and high) and that understanding how the laws affect the middle school population is crucial in determining whether bullying laws work.

It is also important to examine how well the bullying-related outcomes (absent because felt unsafe, in a fight at school, threatened at school, carry weapon at school) measure students' bullying experiences in school because all four correlates may capture delinquency and bullying (i.e., both perpetration and victimization). Therefore, I examine the means of the bullying-related outcomes conditional on being bullied at school. Table 15 displays the raw and

[68] The SCS national means are discussed/shown for 2005-2011 only for the sake of consistency regarding the 'bullied at school' language, see Appendix Figure B-2 which depicts the differences discussed in the main text.
[69] Middle school means are statistically different from high school means at the 5% level. Appendix Table B3 lists the seven types of bullying described in the SCS/NCVS survey.
[70] Means of being electronically bullied at school by another student are statistically different at 1% level.

Table 15. Raw and Conditional Means of Students' Bullying Experiences at School, YRBSS

	1993-2011	2009-2011	
Outcome variable		Conditional on Bullied at school=1	Conditional on Bullied at school=0
Absent because felt unsafe	0.047 (0.211)	0.109 (0.312)	0.032 (0.176)
In a fight at school	0.132 (0.338)	0.183 (0.387)	0.095 (0.294)
Threatened at school	0.074 (0.261)	0.165 (0.371)	0.044 (0.205)
Carried weapon at school	0.071 (0.256)	0.083 (0.276)	0.046 (0.209)
Bullied Index	0.085 (0.278)	0.138 (0.345)	0.052 (0.222)
Electronically bullied[a]	0.163 (0.369)	0.466 (0.499)	0.086 (0.280)
Bullied at school[b]	0.200 (0.400)		

Sample weights provided in the YRBSS are used to calculate the means (standard deviations).
[a] denotes calculations were based on data from YRBSS survey year 2011 only.
[b] denotes calculations were based on data from YRBSS survey years 2009 and 2011 only.

conditional (on whether or not a student was bullied at school) means of the outcome variables used in the YRBSS. It is clear-cut that students who are bullied at school are more likely to be absent, in a fight at school, threatened at school, or to carry a weapon at school than students who are not bullied at school.[71]

Moreover, examination of the pairwise correlations among such outcomes (not shown in main text) reveals the following.[72] First, the pairwise correlations between being bullied at school and the avoidant and violent behaviors are similar to the corresponding conditional

[71] Similar patterns are found in the SCS/NCVS data. Appendix Table B-5 displays conditional means for the SCS/NCVS separately by high (top panel) and middle school (bottom panel). Note that there are two separate columns for SCS/NCVS means because the definition of being bullied at school was expanded after the 2003 survey.

[72] I examine pairwise correlations in both the YRBSS and SCS/NCVS, which are shown in Appendix Tables B-6 – B-8. The SCS/NCVS data is partitioned into two grade level groups—middle and high school—for comparisons to the YRBSS, which surveys high school students. And only the SCS/NCVS data from only survey years 2005-2011 are used because the wording of the bullied question is consistent.

means. Second, being absent, in a fight, or threatened at school have a stronger positive associations with being bullied at school than with carrying a weapon at school.[73] This is reasonable since all four correlates may capture delinquent behavior and both bullying perpetration and victimization, whereas carrying a weapon at school is likely a last resort for most victims of bullying. Third, the bullying index that is made up of all four correlates has a much stronger positive association with carrying a weapon at school (0.655) than with being bullied at school (0.136). There are two main takeaways. The four bullying-related outcomes (absent, fight, threatened, weapon) are reasonable proxies for in-school bullying, however, some of these outcomes (i.e., fight, weapon) might be more akin to bullying perpetration than to bullying victimization (see for example, Nansel et al., 2003). Because these four bullying-related outcomes may also capture, to some degree, delinquent behaviors, one might expect them to be less responsive to states' bullying legislation than outcomes that isolate bullying behaviors such as 'bullied at school'.

Pre-Existing Trends in Student Outcomes

Before turning to the empirical evidence on the impact of states' laws on in-school bullying, I examine pre-existing trends in the four bullying-related outcomes in order to provide some insight about concerns of policy endogeneity. Recall that the DD estimation strategy relies on an implicit assumption that states with any type of bullying legislation (treated states) are on similar trends with states never having bullying legislation put into effect during the sample period and while participating in the YRBSS (control states). I address this issue by estimating

[73] The contextual meaning of threatened at school is different between the YRBSS and SCS/NCVS.The YRBSS asks a student if someone has threatened or injured them with a weapon at school whereas the SCS asks a student whether any student has bullied them—that is, threatened them with harm; see Appendix Tables B2 and B3.

dynamic policy effects, which I discuss later in the empirical results section of this paper. Each graph in Figures 11-14 depicts trends in the national unconditional means of one of the four bullying-related outcomes in the YRBSS. Each graph is centered at states' effective year of "any" type of bullying legislation, which is denoted by 0 on the horizontal axis. 1-2 to right of 0 denotes one to two years after the initial year the bullying law was effective; 3-4 to the right of 0 denotes three to four years after; 5-6 to the right of 0 denotes five to six years after; and 7+ to the right of 0 denotes seven or more years after. Values to the left of 0 on the horizontal axis can be described similarly in terms of years before the bullying law was in effect.

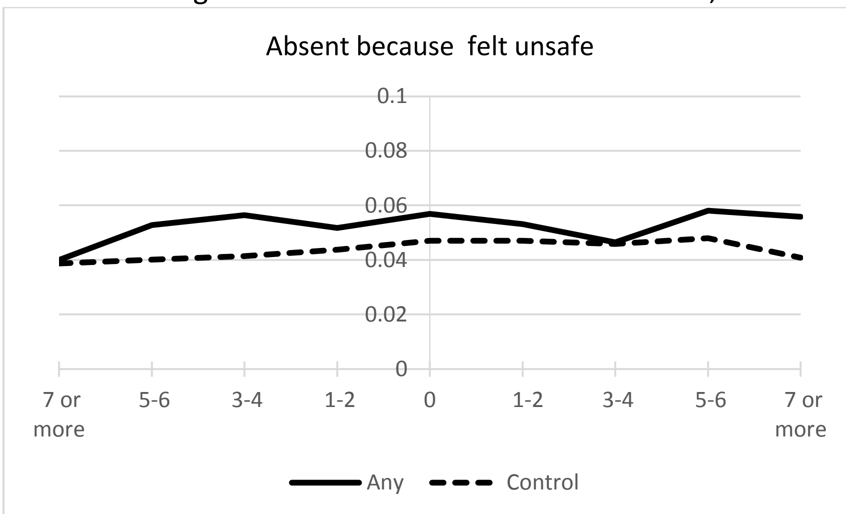
Examining the graphs of pre-existing trends in the four bullying-related outcomes shown in Figures 11-14, overall it appears control and "any" (treated) states are similar in levels, but more important, they have similar trends before "any" type of bullying legislation was in effect (i.e., year 0). 'Absent because felt unsafe' and 'in a physical fight at school' exhibit slightly more variation than 'threatened at school' and 'carried a weapon at school' among the treated states.[74] The unconditional trends for control and "any" (treated) states appear to (mostly) move together after any type of bullying law was put into effect as well. Hence, states' bullying laws appear to be exogenous and there appears to be little descriptive evidence suggesting states' laws reduce in-school bullying among students in grades 9-12.

Empirical Evidence at the Student-level

The top panel of Table 16 presents regression results from estimating "any" type of bullying law effect on each of the four dichotomous measures of bullying-related experiences

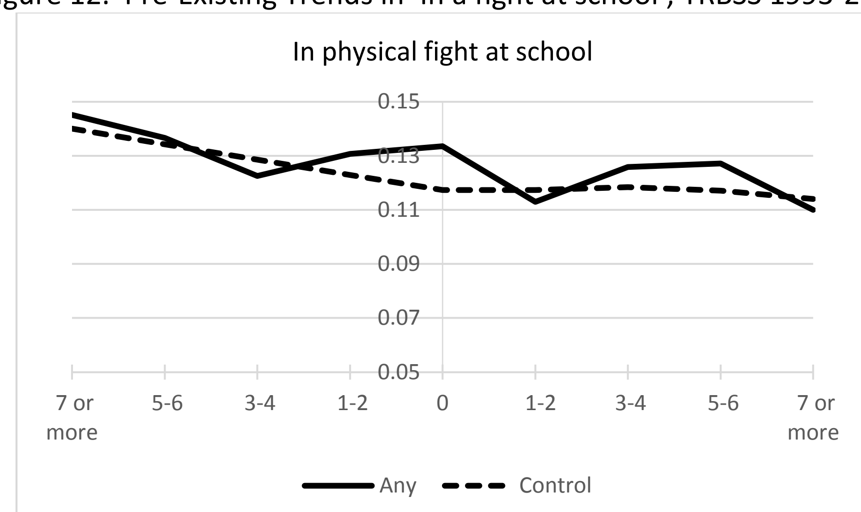
[74] I do not examine pre-existing trends in 'bullied at school' since this question was included only in 2009 and 2011 YRBSS survey years and only two participating states switched.

Figure 11. Pre-Existing Trends in 'Absent because felt unsafe', YRBSS 1993-2011



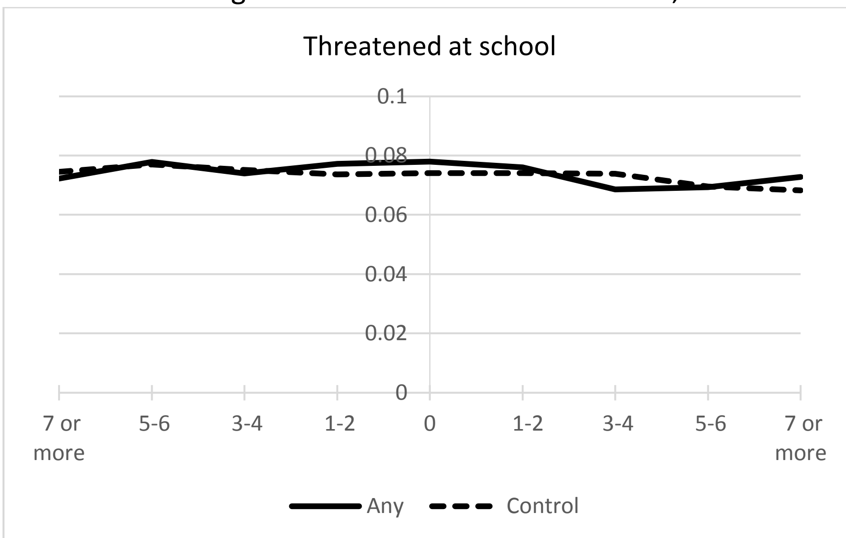
YRBSS sample weights are used.

Figure 12. Pre-Existing Trends in 'In a fight at school', YRBSS 1993-2011



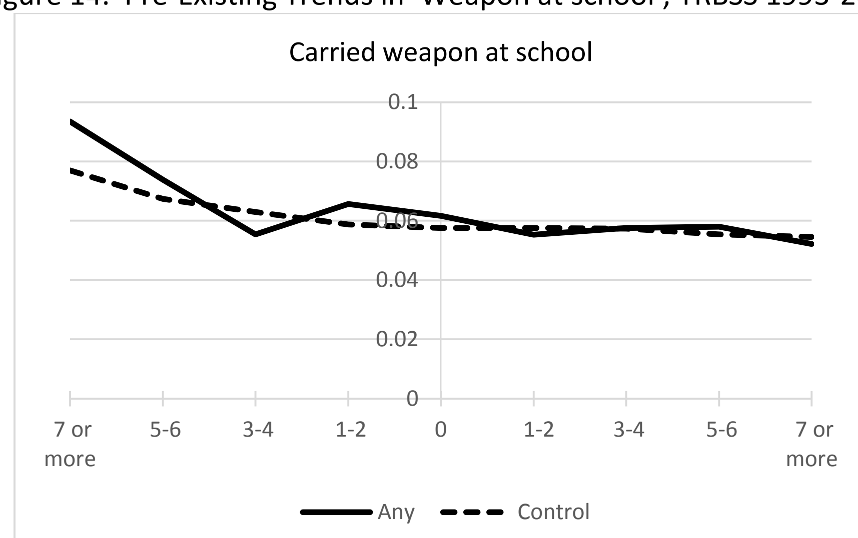
YRBSS sample weights are used.

Figure 13. Pre-Existing Trends in 'Threatened at school', YRBSS 1993-2011



YRBSS sample weights are used.

Figure 14. Pre-Existing Trends in 'Weapon at school', YRBSS 1993-2011



YRBSS sample weights are used.

(column headings indicate outcomes).[75] I estimate that the impact of having "any" type of bullying legislation in effect on 'bullied at school' (not shown) is -0.0167 with standard error of 0.0175, where N = 27,516. But again, the reliability of this result is questionable because it is based on only two states putting bullying laws into effect between 2009 and 2011.[76] Having "any" type of bullying law may reduce the probability a student is absent because they felt unsafe at school by 0.8 percentage points. This result is marginally statistically significant (at the 10% level) and it translates into a 20% reduction in the probability of being absent at least once in the past 30 days, a practically significant impact.[77] The effect of "any" type of bullying law on fighting is small but not statistically significant, reducing the probability of being in a physical fight at school from 13% to 12%. There is no evidence of bullying laws reducing the likelihood of being threatened at school, carrying a weapon at school, or on the bullying index.[78] [79]

[75] I do not show coefficient estimates for control variables since the focus of the analysis is on the policy effects variables and not on correlations between the bullying outcomes and control variables. That said, I note the following. All of the estimated student-level coefficients are statistically and practically significant whereas state-level controls are not significantly different from zero, which is perhaps not surprising given that observations are at the individual level. Student-level coefficients highlight important differences explored in the prior literature on bullying (e.g., Nansel et al., 2001; Wang et al., 2009). Signs and magnitude of coefficient estimates for control variables for other regression results (dichotomous repeated outcome measures, dynamic "any" effects, two-period "weak" and "strong", etc.) are comparable.

[76] And although it is not statistically significant, the magnitude of the impact is practically meaningful; a 1.7 percentage point reduction in the probability of being bullied among students in grades 9-12—a group that is considerably less affected by in-school bullying—means that roughly one-quarter of a million fewer high school students are bullied at school in the past year. Calculations are based on data on total enrollment in grades 9-12 taken from NCES ELSi for academic year 2010-2011.

[77] This calculation is based on a mean of 0.039 for 'absent because felt unsafe' for all students in control states for 1993-2011.

[78] The specifications are also estimated using a probit regression model and the marginal effects are comparable to the OLS estimates presented in Table 16.

[79] Replacing the binary outcomes measures with their frequency measures, which are nearly count variable in nature, I estimate the models using OLS. Signs and statistical significance are unchanged and all estimates become larger in absolute value.

Table 16. DD Estimates of the Effect of Bullying Laws, by Type of Law, YRBSS 1993-2011

Outcomes:	Absent unsafe	In fight at school	Threatened at school	Weapon at school	Bullying index
Policy variable	(1)	(2)	(3)	(4)	(5)
ANY Law	-0.0081*	-0.0102	-0.0022	-0.0006	-0.0020
	(0.0047)	(0.0072)	(0.0047)	(0.0032)	(0.0049)
Observations	133,552	133,552	133,552	133,552	133,552
R-squared	0.0140	0.0434	0.0139	0.0335	0.0273
MANDATE Law	0.0016	0.0032	-0.0001	0.0033	0.0042
	(0.0073)	(0.0115)	(0.0092)	(0.0092)	(0.0110)
Observations	82,208	82,208	82,208	82,208	82,208
R-squared	0.0179	0.0438	0.0146	0.0312	0.0255
MANDATE-DEADLINE Law	-0.0102	0.0034	-0.0105	0.0016	-0.0001
	(0.0085)	(0.0104)	(0.0077)	(0.0055)	(0.0070)
Observations	66,384	66,384	66,384	66,384	66,384
R-squared	0.0121	0.0436	0.0144	0.0374	0.0295

Cluster-robust standard errors in parentheses, clustered by state. *** p<0.01, ** p<0.05, * p<0.10. YRBSS sample weights are used. All regression models control for student demographics (sex, age, race/ethnicity) and grade level and for state-level measures of school quality (per pupil expenditure, pupil-teacher ratio), economic conditions (median household income, unemployment rate), crime (violent crime rate, property crime rate), and include state and year dummies. See text for detailed discussion of outcome and independent variables.

I also investigate how these key results differ without state and/or year fixed effects (results are not shown for reasons of space).[80] I found suggestive evidence that states' laws appear to affect bullying-related behaviors instead of, more broadly, violence in schools.[81] And accounting for both state-specific and year-specific unobservables is important; 40% of the variation in 'absent' and 10% of the variation in 'fight at school' explained by the models is attributed to including both state and year fixed effects in the model.

Next, I estimate the dynamic effects of states' laws on bullying in schools to further investigate policy endogeneity and to study the short-run and long-run policy impacts; these

[80] First, I estimate regression models including only the policy variable and observed student and state characteristics. Then I estimate these same models separately adding in either state or year dummies.
[81] Policy coefficient estimates are essentially zero when only year fixed effects are introduced. While not statistically significant, all policy coefficient estimates are negative but small in magnitude when only state fixed effects are introduced.

results are shown in Table 17. The lagged policy variables allow one to empirically examine whether the estimated impacts of states' bullying legislation are driven by policy endogeneity. The lead policy variables allow one to examine whether the estimated impacts of states' bullying legislation change over time. It is worthwhile to note that the standard errors of the policy variables are larger for the dynamic effects (1-2 years after, 3-4 years after, 5-6 years after, etc.) than for the pre/post effects (Table 16), which one might expect since states' YRBSS participation fluctuates each survey year and the "any" bullying law effect now has a time series dimension. Overall, the pattern is very small positive coefficients before the effective year and small negative coefficients after the effective where the majority of estimates are not statistically significant. This empirical evidence coupled with the graphs of pre-existing trends in unconditional means (Figures 11-14) provides support that adoption of bullying laws is likely exogenous. In other words, it is reasonable to assume that states' decisions to enact bullying legislation are uncorrelated with unobserved time-varying state-specific characteristics that are also correlated with in-school bullying outcomes. Additionally, the lead policy effects appear to become increasingly more negative over time, particularly after three years. This provides some suggestive evidence that it may take some time (possibly three years more) for state bullying laws to have an impact in schools.

Another consideration deals with the repeated nature of bullying and whether estimates based on binary forms of the proxies for in-school bullying capture this aspect. To address this important concern, I estimate the two-period policy effects on 'repeated'

Table 17. Estimates of the Dynamic Effects of ANY Bullying Laws, YRBSS 1993-2011

Outcomes: Policy variables	Absent unsafe (1)	In fight at school (2)	Threatened at school (3)	Weapon at school (4)	Bullying index (5)
7+ years before	0.0066	0.0103	0.0026	0.0035	0.0124
	(0.0076)	(0.0103)	(0.0073)	(0.0081)	(0.0081)
5-6 years before	0.0069	0.0034	0.0057	-0.0018	0.0087
	(0.0072)	(0.0093)	(0.0057)	(0.0074)	(0.0058)
3-4 years before	0.0047	-0.0057	0.0021	-0.0068	0.0052
	(0.0056)	(0.0085)	(0.0052)	(0.0071)	(0.0067)
Year effective	0.0031	-0.0088	0.0034	-0.0055	0.0079
	(0.0077)	(0.0143)	(0.0066)	(0.0081)	(0.0126)
1-2 years after	-0.0039	-0.0141	0.0026	-0.0023	0.0008
	(0.0078)	(0.0118)	(0.0072)	(0.0062)	(0.0078)
3-4 years after	-0.0135**	-0.0105	-0.0077	-0.0030	-0.0031
	(0.0055)	(0.0117)	(0.0076)	(0.0070)	(0.0068)
5-6 years after	-0.0120	-0.0094	-0.0083	-0.0088	-0.0045
	(0.0075)	(0.0104)	(0.0073)	(0.0083)	(0.0096)
7+ years after	-0.0115	-0.0168	-0.0052	-0.0133**	-0.0038
	(0.0085)	(0.0136)	(0.0065)	(0.0058)	(0.0065)
Observations	133,552	133,552	133,552	133,552	133,552
R-squared	0.0141	0.0435	0.0139	0.0336	0.0274

Cluster-robust standard errors in parentheses, clustered by state. *** p<0.01, ** p<0.05, * p<0.10. YRBSS sample weights are used. 1-2 years before ANY Bullying Law is the omitted policy category. All regression models control for student demographics (sex, age, race/ethnicity) and grade level and for state-level measures of school quality (per pupil expenditure, pupil-teacher ratio), economic conditions (median household income, unemployment rate), crime (violent crime rate, property crime rate), and include state and year dummies. See text for detailed discussion of outcome and independent variables.

measures of the four (binary) bullying-related outcomes.[82] These results (shown in Appendix Table B9) lead to an examination of whether the impact of states' laws varies across the frequency distribution of the bullying-related outcomes.[83] It is plausible that states' laws may reduce how often (intensive margin) a student experiences bullying-related behaviors without affecting whether or not s/he responds affirmatively to having had experienced the behavior at

[82] Recall from YRBSS data section that the dependent variable equals one if a student experiences the bullying-related outcome two or more days/times in the previous period.

[83] Signs and statistical significance for all bullying proxies are unaffected by recoding the outcomes even though the standard errors are slightly smaller. In contrast to the magnitudes of 'repeatedly threatened at school' and 'repeatedly carried a weapon at school', which do not change, the magnitudes of the coefficients on 'repeatedly absent' and 'repeatedly in a fight at school' are reduced to roughly one-half and one-third, respectively, of their non-repeated binary counterparts.

all in the previous period (extensive margin).[84] Such estimates (Appendix Figures B3 – B6), however, suggest the impact of having "any" type of bullying law in effect is the greatest at the extensive margin for absenteeism only, all else held constant.[85]

At the outset of this paper, I discussed variation across states' legislation as one motivation for undertaking this study and for the identification strategy used in the evaluation. Thus far I have exploited variation in the timing of states' effective year of legislation to estimate the impact of bullying laws. Next, I exploit variation in whether states' legislation penetrates to the district in order to estimate the impact of different types of bullying laws. States' bullying laws can also be classified into mutually exclusive categories "mandate" or "mandate-deadline" based on a rationale discussed earlier. To briefly recap, "mandate" types of bullying laws only require that school districts develop and implement local policies on bullying—as opposed to laws without such a mandate, whereas "mandate-deadline" types of bullying laws require that school districts develop and implement local policies on bullying and set a specific date by which districts must comply with this mandate. One would expect the "mandate-deadline" type of laws to have a larger impact on in-school bullying than the "mandate" type of laws, all else equal.

Empirical results from estimating equation (1) by OLS for "mandate" and "mandate-deadline" laws are presented in the top and bottom portions of Table 16, respectively. Dependent variables (denoted by column headings) consist of dichotomous measures of

[84] To investigate this, I estimated the marginal effects of "any" type of bullying legislation on the predicted probability of each bullying-related experience at each frequency category. Marginal effects are predicted based on a 14 year-old, white, male student in grade 9 at the mean values of the state-level covariates.
[85] All marginal effects are statistically significant at the 10% level for 'absent because felt unsafe'. Marginal effects for 'in a fight at school', 'threated at school', and 'carried a weapon at school' are not significant even at the 10% level.

bullying related outcomes.[86] Note that the standard errors for both “mandate” and “mandate-deadline” policy variables are roughly one-third larger than those for “any” policy variables, which is not surprising given that there are roughly an equal number of states have a “mandate” and “mandate-deadline” type of bullying law whereas there are slight more than twice as many states have “any” type of bullying law.

The “mandate” bullying law effect on ‘being bullied at school’ (not shown) is -0.1248 with standard error 0.2358 where N = 17,944, which is based on only one state putting a “mandate” bullying law into effect. Yet there is no other (reliable) evidence that “mandate” bullying laws reduce (or increase) bullying-related outcomes in schools. The “mandate-deadline” bullying law effect on ‘being bullied at school’ (not shown) is 0.0129 with standard error 0.0277 where N = 12,046, which is also based on only one state putting a “mandate-deadline” bullying law into effect. The “mandate-deadline” law effects are rather mixed in terms of signs and magnitude. Instead of seeing across the board stronger negative effects for the “mandate-deadline” estimates as compared to the “mandate” estimates, larger negative effects arise for ‘absent’ and ‘threatened’ (both are statistically insignificant).

To summarize, the student-level empirical results suggest there is little evidence supporting that states’ bullying laws may reduce bullying in high schools across the U.S. Of the reliable estimates presented here, states’ having any type of bullying legislation in effect reduces the probability of a student being absent at least once in the past month because s/he felt unsafe at school by 20% or 0.8 percentage points (at 10% level).

[86] Results for their “repeated” counterparts are shown in the middle and bottom portions of Appendix Table B9 for “mandate” and “mandate-deadline”, respectively.

Descriptive Evidence at the School-level

Table 18 displays a frequency table of student bullying by year for the weighted SSOCS sample, which consists of all grade levels. The distribution of principals' reports of how often student bullying occurs at their school is mostly unchanged over time. Roughly 17% consistently report student bullying occurs at least once a week; more than half consistently report student bullying occurs on occasion; and only 2% consistently report student bullying never occurs at their school. Yet there does appear to be a downward trend from 9.9% in 2004 to 6.9% in 2010 for the 'happens daily' category and a upward trend for the 'at least once a month' category from 19.5% in 2004 to 22.5% in 2010 indicating that the frequency of in-school bullying might be subsiding over time. Sample means of bullying-related outcomes by year are shown in Table 19. ADA rates are stable at 94% and weapon usage/possession rates hover at 0.002. Fight rates exhibit a downward trend from 0.027 in 2004 to 0.018 in 2010 and the net total mobility rate (i.e., Transfers-IN rate less transfers-OUT rate) increases from 0.005 in 2004 to 0.030 in 2010.

Since the SSOCS surveys schools consisting of all grade levels and since middle school students are more likely to report being bullied at school than high school students (recall descriptive evidence from SCS/NCVS), I examine school-level measures of the frequency of student bullying by grade level as shown in Figure 15.[87] The frequency distribution for middle school principals' reports of how often student bullying occurs is markedly shifted to the left of all other grade level distributions and there is little difference between the elementary and high school distributions. The descriptive evidence is clear-cut that student bullying occurs much more often in middle school than in either elementary or high school. This result is especially

[87] Appendix Table B4 shows the composition of the SSOCS sample (weighted and unweighted) by grade level.

Table 18. Frequency of Bullying Over Time, SSOCS

How often student bullying occurs at your school	2004	2006	2008	2010
Never happens	1.7%	2.1%	2.0%	2.6%
Happens on occasion	51.8%	55.1%	52.0%	51.8%
At least once a month	19.5%	18.3%	20.5%	22.5%
At least once a week	17.2%	17.8%	17.0%	16.2%
Happens daily	9.9%	6.7%	8.5%	6.9%

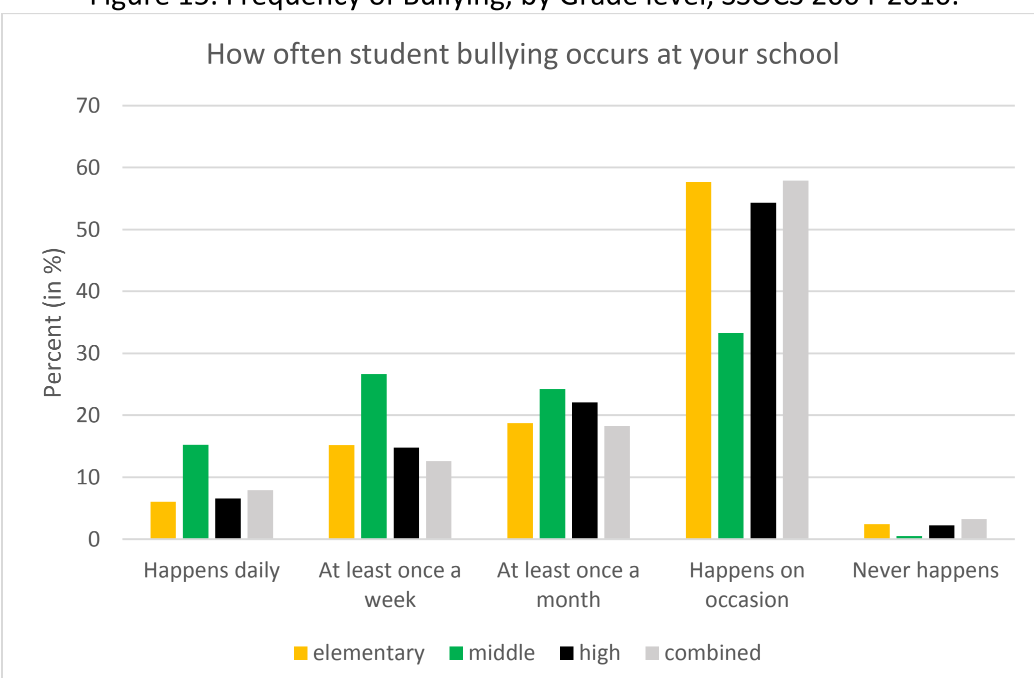
SSOCS sampling weights are used. All grade levels combined. See text for discussion of outcomes.

Table 19. Means of Bullying-related Outcomes Over Time, SSOCS

Outcome variable	2004	2006	2008	2010
Average Daily Attendance (ADA) rate	0.938	0.936	0.938	0.938
Total number of Transfers-IN to the school as a share of total enrollment	0.091	0.105	0.099	0.105
Total number of Transfers-OUT of the school as a share of total enrollment	0.086	0.091	0.083	0.075
Total number of recorded physical attacks/fights as a share of total student enrollment	0.027	0.022	0.024	0.018
Total number of recorded weapon usages/possessions as a share of total student enrollment	0.002	0.003	0.001	0.002

SSOCS sampling weights are used. All grade levels combined. See text for discussion of outcomes.

Figure 15. Frequency of Bullying, by Grade level, SSOCS 2004-2010.



SSOCS sampling weights are used. See text for discussion of grade levels and outcomes.

striking considering that this is a holistic measure based on subjective accounts by principals and that some incidents of student bullying are likely never reported to administration.

To examine the relationships between bullying and avoidant or violent behaviors in the SSOCS school-level data, pairwise correlations by grade level are presented in Tables 20-23.[88] Fight rate has the strongest positive association with bullying and the relationship becomes stronger as the highest grade level in the school decreases (i.e., moving from high to elementary school). The IN-transfers rates and the OUT-transfers rates have strong positive correlations with each other. And while their relationships with bullying are somewhat mixed, each pairwise correlation becomes stronger and more positive as the highest grade level in the school decreases (i.e., moving from high to elementary school). Not surprisingly, given that ADA rates exhibit little variation over time, the magnitude of the correlation between ADA rate and the frequency of bullying is close to zero across all levels even though its sign is always negative. A similar statement can be made about weapon rates except its sign is mostly positive. Overall these associations align well with those found in the student-level data.

There are two main takeaways from these results. Findings from the analysis of school-level descriptive evidence are consistent with the findings from the analysis of student-level descriptive evidence; bullying-related outcomes such as avoidant and violent behaviors serve as reasonable and reliable proxies for student bullying. Furthermore, the relationships between bullying and bullying- related outcomes vary by grade level in such a way that empirical estimates from analyses based on data on high school student populations may indeed serve as

[88] Recall from the SSOCS data description that these measures are based on recorded incidents—not holistic assessments, however, as already discussed above administration data may underestimate actual incidents.

Table 20. Pairwise Correlations, High school, SSOCS 2004-2010

	Freq. bullying	ADA rate	IN-transfers rate	OUT-transfers rate	Fight rate	Weapon rate
Frequency of bullying	1.000					
ADA rate	-0.007	1.000				
IN-transfer rate	-0.038	-0.182	1.000			
OUT-transfer rate	0.017	-0.136	0.620	1.000		
Fight rate	0.093	-0.093	0.157	0.226	1.000	
Weapon rate	0.022	-0.007	0.038	0.041	0.058	1.000

SSOCS sampling weights are used. High school refers to grades 9-12. See text for discussion of outcomes.

Table 21. Pairwise Correlations, Middle school, SSOCS 2004-2010

	Freq. bullying	ADA rate	IN-transfers rate	OUT-transfers rate	Fight rate	Weapon rate
Frequency of bullying	1.000					
ADA rate	-0.015	1.000				
IN-transfer rate	0.071	-0.092	1.000			
OUT-transfer rate	0.067	-0.062	0.604	1.000		
Fight rate	0.144	-0.090	0.277	0.299	1.000	
Weapon rate	-0.006	-0.009	0.020	0.013	0.037	1.000

SSOCS sampling weights are used. Middle school refers to grades 6-8. See text for discussion of outcomes.

Table 22. Pairwise Correlations, Elementary school, SSOCS 2004-2010

	Freq. bullying	ADA rate	IN-transfers rate	OUT-transfers rate	Fight rate	Weapon rate
Frequency of bullying	1.000					
ADA rate	-0.006	1.000				
IN-transfer rate	0.101	-0.019	1.000			
OUT-transfer rate	0.145	-0.050	0.705	1.000		
Fight rate	0.201	-0.028	0.145	0.196	1.000	
Weapon rate	0.008	0.000	0.062	0.048	0.044	1.000

SSOCS sampling weights are used. Elementary school refers to grades K-5. See text for discussion of outcomes.

Table 23. Pairwise Correlations, Combined, SSOCS 2004-2010

	Freq. bullying	ADA rate	IN-transfers rate	OUT-transfers rate	Fight rate	Weapon rate
Frequency of bullying	1.000					
ADA rate	0.035	1.000				
IN-transfer rate	-0.063	-0.158	1.000			
OUT-transfer rate	-0.039	-0.256	0.152	1.000		
Fight rate	0.102	0.005	0.110	0.445	1.000	
Weapon rate	0.037	-0.121	0.004	0.085	0.010	1.000

SSOCS sampling weights are used. See text for discussion of combined grade levels and outcomes.

lower bounds for empirical estimates from analyses based on data on middle school or on elementary school student populations.

Empirical Evidence at the School-level

Marginal effects of having “any” bullying legislation on the predicted probability of each frequency category of student bullying are shown in Table 24. The laws have no effect on how often student bullying occurs at the either the combined grade, middle, or high school levels, however, the laws appear to have an effect at the elementary school level. Interestingly, having any bullying law in effect increases the probabilities that a principal in an elementary school reports student bullying occurs at least once a month, at least once a week, and almost daily and decreases the probabilities that the principal reports student bullying occurs on occasion or never happens.

Also presented in Table 24 are the marginal effects of a being an elementary, middle, or combined school as compared to being a high school. This empirical evidence confirms two relationships appearing in the descriptive evidence. First, bullying occurs more frequently in middle school as compared to high school, holding all else constant. Second, there are no differences between elementary and high school or between schools with combined grades and high school across the frequency categories of student bullying, all else the same.

Table 25 presents regression results from estimating the “any” bullying law effect by grade level group on each of the five school-level measures of bullying-related outcomes, where the dependent variables are denoted by column headings.[89] There is little to no

[89] Coefficient estimates for control variables are not shown. Most of the estimated coefficients for the school characteristics are neither practically nor statistically significant in the specifications; three exceptions are the

Table 24. Marginal Effects of ANY Bullying Laws on Frequency of Bullying, by Grade level, SSOCS 2004-2010

Frequency: Covariates	Never happens (1)	Happens on occasion (2)	At least once a month (3)	At least once a week (4)	Happens daily (5)
ANY Law	0.0007	0.0035	-0.0013	-0.0018	-0.0011
	(0.0046)	(0.0217)	(0.0079)	(0.0115)	(0.0070)
ANY Law X Middle	-0.0036	-0.0178	0.0062	0.0093	0.0058
	(0.0041)	(0.0221)	(0.0072)	(0.0115)	(0.0075)
ANY Law X Elementary	-0.0140***	-0.0712***	0.0243***	0.0371***	0.0237***
	(0.0033)	(0.0221)	(0.0055)	(0.0107)	(0.0084)
ANY Law X Combined	-0.0026	-0.0128	0.0045	0.0067	0.0041
	(0.0090)	(0.0471)	(0.0160)	(0.0245)	(0.0155)
Middle	-0.0237***	-0.2365***	0.0387***	0.1116***	0.1099***
	(0.0031)	(0.0153)	(0.0056)	(0.0077)	(0.0075)
Elementary	0.0003	0.0016	-0.0006	-0.0008	-0.0005
	(0.0040)	(0.0189)	(0.0069)	(0.0100)	(0.0060)
Combined	-0.0037	-0.0189	0.0065	0.0099	0.0062
	(0.0069)	(0.0376)	(0.0125)	(0.0195)	(0.0125)
Observations = 10,520[a]					
Pseudo R^2 = 0.0467					

[a] Unweighted sample entities are rounded to the nearest ten per IES publication policy.
Cluster-robust standard errors in parentheses, clustered by state. *** p<0.01, ** p<0.05, * p<0.10. SSOCS sampling weights are used. Ordered probit regression model includes the same independent variables as the linear regression models estimated in Table 13. Elementary refers to grades K-5; middle refers to grades 6-8; high school refers to grades 9-12; and combined refers to schools with combined grade levels. Marginal effects are predicted for high schools located in a low crime area at the mean values of all other covariates. See text for discussion of outcome and independent variables.

---

percent of black students, the percent of students eligible for free or reduced lunch, and (for 'fight' and 'transfers-OUT') the percent of Hispanic students.

Table 25. DD Estimates of Effect of ANY Bullying Laws, by Grade Level, SSOCS 2004-2010

Outcomes:	ADA	IN	OUT	fight	weapon
Independent variables	(1)	(2)	(3)	(4)	(5)
ANY Bullying Law	0.0026	0.0151	-0.0015	0.0066*	-0.0010
	(0.0049)	(0.0171)	(0.0082)	(0.0038)	(0.0009)
ANY Law X Middle	-0.0037	-0.0217	-0.0020	-0.0040	0.0020
	(0.0040)	(0.0174)	(0.0075)	(0.0028)	(0.0013)
ANY Law X Elementary	-0.0029	-0.0173	-0.0040	-0.0013	0.0017
	(0.0054)	(0.0186)	(0.0069)	(0.0034)	(0.0011)
ANY Law X Combined	-0.0139**	-0.0415	0.0246	0.0200	0.0011
	(0.0066)	(0.0644)	(0.0152)	(0.0191)	(0.0011)
Middle	0.0195***	-0.0065	-0.0174***	0.0151***	-0.0015
	(0.0036)	(0.0099)	(0.0056)	(0.0029)	(0.0010)
Elementary	0.0214***	-0.0006	-0.0173***	-0.0066*	-0.0027***
	(0.0047)	(0.0103)	(0.0064)	(0.0038)	(0.0010)
Combined	0.0190***	0.0526	-0.0290**	-0.0069***	-0.0023*
	(0.0041)	(0.0707)	(0.0130)	(0.0021)	(0.0012)
Observations[a]	10,520	10,520	10,520	10,520	10,520
R-squared	0.0369	0.1463	0.1385	0.0742	0.0106

[a] Unweighted sample entities are rounded to the nearest ten per IES publication policy.

Cluster-robust standard errors in parentheses, clustered by state. *** p<0.01, ** p<0.05, * p<0.10. SSOCS sampling weights are used. All regression models control for school demographics (% female, %black, %Hispanic, % other race), grade level group, school quality (per pupil expenditure, pupil-teacher ratio), economic conditions (% eligible for free/reduced lunch), local crime level, and include state and year dummies. Elementary refers to grades K-5; middle refers to grades 6-8; high school refers to grades 9-12; and combined refers to schools with combined grade levels. See text for discussion of outcome and independent variables.

evidence of “any” type of bullying laws having an effect on any of the five bullying-related outcomes at any grade level group. Two policy effects are significant at the 5% level: ‘fight rate’ in high schools and ‘ADA rate’ in schools with combined grades.[90] Since there are 20 estimated policy effects, one of these one could be statistically significant by chance at the 5% level.[91]

To investigate whether the impact of the laws may differ by type of legislation, I estimate separately the policy effects by grade level group for “mandate” and “mandate-deadline” laws using the same control group. These results are shown in the top and bottom panels Table 26, respectively. Having either a “mandate” or “mandate-deadline” bullying law in effect increases the probabilities of student bullying occurring at least once a month, at least once a week, or daily, and decreases the probabilities of student bullying occurring on occasion or never at all, all else the same, in elementary schools. A similar pattern in the “mandate” law effects is evident for middle schools; however, “mandate-deadline” laws have no effect on how often student bullying occurs in middle schools. Neither type of law has an effect on the frequency of student bullying in combined grade or in high schools. These results could be explained by “mandate” laws failing to provide a strict deadline by which schools must comply with state mandates to take actions against bullying in schools (i.e., credible threats to bullies). Arguably, putting a bullying law into effect likely increases awareness about bullying among the students. Then one could see students’ reporting more incidents of bullying to administration after a putting into effect a law that lacks a credible threat to bullies even though actual

[90] Typically schools with combined grade levels are either 7-12 or K-12.
[91] After the dissertation, I plan to perform a falsification test on the SSOCS results.

Table 26. Marginal Effects of MANDATE Bullying Laws on Frequency of Bullying, by Grade level SSOCS 2004-2010

Covariates — Frequency:	Never happens (1)	Happens on occasion (2)	At least once a month (3)	At least once a week (4)	Happens daily (5)
MANDATE Law	0.0044	0.0211	-0.0076	-0.0113	-0.0067
	(0.0061)	(0.0293)	(0.0103)	(0.0156)	(0.0096)
MANDATE Law X Middle	-0.0111**	-0.0649**	0.0201***	0.0337**	0.0221*
	(0.0041)	(0.0301)	(0.0072)	(0.0153)	(0.0119)
MANDATE Law X Elementary	-0.0181***	-0.0973***	0.0316***	0.0509***	0.0329***
	(0.0038)	(0.0237)	(0.0052)	(0.0119)	(0.0106)
MANDATE Law X Combined	-0.0011	-0.0052	0.0018	0.0028	0.0016
	(0.0138)	(0.0689)	(0.0241)	(0.0366)	(0.0219)
Middle	-0.0228***	-0.2273***	0.0385***	0.1089***	0.1026***
	(0.0032)	(0.0214)	(0.0072)	(0.0104)	(0.0100)
Elementary	0.0010	0.0049	-0.0018	-0.0026	-0.0015
	(0.0052)	(0.0242)	(0.0088)	(0.0130)	(0.0076)
Combined	-0.0062	-0.0341	0.0111	0.0179	0.0113
	(0.0096)	(0.0592)	(0.0177)	(0.0306)	(0.0206)
Observations = 5,310[a]					
Pseudo R^2 = 0.0488					
MANDATE-DEADLINE (MD) Law	-0.0042	-0.0168	0.0060	0.0094	0.0056
	(0.0066)	(0.0271)	(0.0096)	(0.0149)	(0.0092)
MD LAW X Middle	-0.0005	-0.0018	0.0007	0.0010	0.0060
	(0.0068)	(0.0269)	(0.0097)	(0.0151)	(0.0090)
MD Law X Elementary	-0.0128***	-0.0566**	0.0188***	0.0310**	0.0196*
	(0.0051)	(0.0289)	(0.0075)	(0.0151)	(0.0112)
MD Law X Combined	-0.0009	-0.0034	0.0012	0.0019	0.0011
	(0.0157)	(0.0636)	(0.0226)	(0.0354)	(0.0213)
Middle	-0.0279***	-0.2303***	0.0374***	0.1121***	0.1086***
	(0.0051)	(0.0186)	(0.0063)	(0.0103)	(0.0104)
Elementary	0.0003	0.0012	-0.0004	-0.0007	-0.0004
	(0.0056)	(0.0217)	(0.0079)	(0.0121)	(0.0072)
Combined	-0.0078	-0.0353	0.0115	0.0192	0.0122
	(0.0098)	(0.0500)	(0.0149)	(0.0268)	(0.0181)
Observations = 5,200a					
Pseudo R2 = 0.0485					

[a] Unweighted sample entities are rounded to the nearest ten per IES publication policy.
Cluster-robust standard errors in parentheses, clustered by state. *** p<0.01, ** p<0.05, * p<0.10. SSOCS sampling weights are used. Ordered probit regression model includes the same independent variables as the linear regression models estimated in Table 13. Elementary refers to grades K-5; middle refers to grades 6-8; high school refers to grades 9-12; and combined refers to schools with combined grade levels. Marginal effects are predicted for high schools located in a low crime area at the mean values of all other covariates. See text for discussion of outcome and independent variables.

incidents of student bullying may not have changed.[92] In fact, the “mandate-deadline” policy effects on student bullying frequency are about 2 percentage points smaller in absolute value than the “mandate” policy effects. Coupling this with evidence (here and in the prior literature) showing bullying reports decrease with age may explain the policy effects by grade level groups.

The top and bottom panels of Table 27 present estimates of the “mandate” and “mandate-deadline” policy effects on avoidant and violent behaviors and transfer rates (or bullying-related outcomes). “Mandate” bullying laws have no effect on nearly all bullying-related outcomes at any grade level with one exception--‘fight rate’ in high schools. Yet one of the twenty estimated policy effects could be significant by chance at the 5% level. “Mandate-deadline” bullying laws have no effect on fight, weapon, or transfer rates, however, these laws affect ADA rates differently across grade level groups. ADA rates increase by 0.0183 (at the 1% level of significance) in high schools after a “mandate-deadline” law is in effect. In contrast, ADA rates decrease by 0.0086 in middle schools and by 0.0209 in combined grade schools (both at the 10% level) after a “mandate-deadline” law goes into effect. Again, it is possible that one of these policy effects is due to chance. That said, the results are difficult to explain given that combined schools typically consist of grades 7-12 or K-12 and only make up about 4% of the schools in the sample.

The principle takeaways are that bullying occurs more frequently in middle school than in high school and the type of bullying law may matter, all else constant. School-level analysis

[92] Arguably, one might explain that a consequence of states’ laws is an hyper-sensitivity to bullying, especially among younger students, and thus an over-reporting occurs. I assume principals reports reflect on accurate incidents of student bullying in their schools—not merely allegations of bullying.

Table 27. DD Estimates of Effect of MANDATE Bullying Laws, by Grade Level, SSOCS 2004-2010

Outcomes:	ADA	IN	OUT	fight	weapon
Covariates	(1)	(2)	(3)	(4)	(5)
MANDATE Bullying Law	0.0054	0.0264	-0.0000	0.0099*	-0.0012
	(0.0061)	(0.0349)	(0.0146)	(0.0055)	(0.0009)
MANDATE Law X Middle	-0.0012	-0.0307	-0.0067	-0.0061	0.0015
	(0.0067)	(0.0342)	(0.0131)	(0.0041)	(0.0013)
MANDATE Law X Elementary	-0.0020	-0.0377	-0.0087	-0.0047	0.0010
	(0.0087)	(0.0377)	(0.0125)	(0.0047)	(0.0014)
MANDATE Law X Combined	-0.0146	-0.1110	0.0195	0.0033	0.0021
	(0.0117)	(0.1046)	(0.0250)	(0.0072)	(0.0016)
Middle	0.0232***	-0.0186	-0.0241**	0.0156***	-0.0020
	(0.0051)	(0.0178)	(0.0093)	(0.0035)	(0.0014)
Elementary	0.0271***	-0.0097	-0.0265**	-0.0051	-0.0028*
	(0.0065)	(0.0174)	(0.0111)	(0.0036)	(0.0015)
Combined	0.0195***	0.0979	-0.0331	-0.0079***	-0.0025
	(0.0054)	(0.1126)	(0.0223)	(0.0017)	(0.0016)
Observations[a]	5,310	5,310	5,310	5,310	5,310
R-squared	0.0494	0.1618	0.1221	0.1255	0.0185
MANDATE-DEADLINE (MD) Law	0.0183***	0.0057	-0.0061	-0.0033	0.0003
	(0.0061)	(0.0123)	(0.0071)	(0.0054)	(0.0008)
MD Law X Middle	-0.0086*	-0.0109	0.0079	0.0025	0.0024
	(0.0048)	(0.0118)	(0.0074)	(0.0039)	(0.0025)
MD Law X Elementary	-0.0050	-0.0028	0.0058	0.0050	0.0004
	(0.0068)	(0.0106)	(0.0054)	(0.0048)	(0.0011)
MD Law X Combined	-0.0209**	-0.1045	0.0123	-0.0253	-0.0007
	(0.0093)	(0.0922)	(0.0157)	(0.0294)	(0.0012)
Middle	0.0194***	0.0053	-0.0090*	0.0114***	-0.0001
	(0.0042)	(0.0119)	(0.0046)	(0.0037)	(0.0007)
Elementary	0.0217***	0.0143	-0.0062	-0.0093	-0.0015***
	(0.0059)	(0.0102)	(0.0049)	(0.0069)	(0.0005)
Combined	0.0213***	0.1151	-0.0168	0.0192	-0.0010
	(0.0052)	(0.0983)	(0.0108)	(0.0258)	(0.0011)
Observations[a]	5,200	5,200	5,200	5,200	5,200
R-squared	0.0516	0.1651	0.1538	0.0696	0.0135

[a] Unweighted sample entities are rounded to the nearest ten per IES publication policy.
Cluster-robust standard errors in parentheses, clustered by state. *** p<0.01, ** p<0.05, * p<0.10. SSOCS sampling weights are used. Same regression model as in Table 13. Elementary refers to grades K-5; middle refers to grades 6-8; high school refers to grades 9-12; and combined refers to schools with combined grade levels. See text for discussion of outcome and independent variables.

.

also suggests bullying laws have little to no effect on student bullying in high school, which aligns with results from the student-level analysis. Lastly, there is suggestive evidence that the laws may increase student bullying in middle and elementary schools, which may be explained by a heightened awareness of what constitutes bullying behavior and thus increased reporting of bullying incidents to school administrators.

Conclusion

Recent media attention has stressed the prevalence and persistence of bullying among youths. Researchers have studied correlates of bullying and have provided legislative recommendations to address bullying in U.S. schools. Examining whether, and to what extent, state laws reduce bullying in schools is important for informing decisions regarding public policy at both state and federal levels. From a purely economic standpoint, bullying can be financially problematic for schools; bullying might have serious negative labor market consequences if it impedes human capital accumulation; and bullying during youth may carry over into adulthood and have longer-run impacts within households, labor markets, and the larger economy.

I evaluate whether bullying legislation abates bullying in schools by employing a difference-in-differences approach that exploits variation across states in the timing and type of law adopted in conjunction with nationally representative surveys of students' bullying-related experiences in school. My main results suggest that state laws may have little impact at reducing bullying in high school. Specifically, I find that having any type of bullying law in effect decreases the probability that a high school student is absent at least once in the past month because s/he felt unsafe at school by 20% (at the 10% level). I find the laws have no significant effect on other bullying-related outcomes in schools such as being in a fight, being threatened, or carrying a weapon.

To dispel concerns about policy endogeneity, I present figures showing pre-existing trends in bullying outcomes and estimates of lagged policy effects. The evidence from both descriptive and empirical approaches supports the assumption that states’ decision to implement any type of bullying legislation is exogenous. Estimates of short-run and long-run policy effects provide suggestive evidence that the laws may become more effective at reducing bullying over time, particularly after three years. While the laws appear to have little to no effect at the extensive margin across multiple measures of bullying, it is possible the laws could have an effect at the intensive margin. So I also utilize frequency measures of the bullying outcomes in the student-level analysis, revealing that having any type of bullying law in effect has the greatest impact at the extensive margin for whether a student is absent because s/he felt unsafe at school; the laws appear to have no impact at either the extensive or intensive margin for being in a fight at school, threatened at school, or for carrying a weapon at school.

Lastly, I examine whether the type of bullying legislation matters by classifying states’ laws by the degree to which it potentially filters down to the school district level. That is, states’ with bullying laws in effect may have “mandate” bullying laws or they may have “mandate-deadline” laws. Keeping the control states consistent with those used in the main analysis, which examines having “any” type of bullying legislation, allows me to compare estimates of the impact of each type of legislation. While not statistically significant possibly due to partitioning the sample, I find some empirical support for the notion that those laws specifying a date by which school districts need to comply with state mandates may be more effective at reducing bullying in schools than laws with state mandates but no deadline.

Overall the student-level analysis provides little evidence that states' laws reduce in-school bullying although caveats exist. Data limitations in the large nationally representative student-level survey require one to use outcomes that likely measure both delinquency and bullying, which might produce underestimates of the legislative impact on bullying experiences among high school students.

The school-level analysis allows me to examine the external validity of results from the main (student-level) analysis. I find that the estimated impacts of bullying laws on high school students likely do not extend to middle or elementary school populations. Additional challenges to evaluating state bully legislation are revealed. Student bullying occurs most frequently in middle schools yet data on this population lacks state-identifiers. Different types of laws (i.e., "mandate", "mandate-deadline") may have heterogeneous effects across elementary, middle, and high schools. Finally, there is suggestive evidence that the laws may increase reports of student bullying, especially among elementary students—a population that is not included in current large-scale nationally surveys on of students experiences in schools.

How can future work build on the contributions of this study? After completion of the dissertation, I plan to perform falsification tests of the estimated impacts of state bullying legislation from the school-level analysis since it does not make sense to estimate dynamic policy effects here. Second, I am considering updating the evaluation by including the next waves of the biennial survey data (i.e., YRBSS 2013) and by updating my data on state bullying laws accordingly. Third, I am also considering investigating whether controlling for school policies and programs changes the estimated impacts of bullying laws in the school-level analysis. Lastly, I plan to investigate how one might work with the appropriate agencies in order to gain access to electronic forms of the SCS/NCVS that include state-identifiers going back to the early 1990's. The SCS/NCVS has some important advantages over the YRBSS. Utilizing SCS/NCVS student-level data that includes state identifiers would improve the evaluation of states' bully laws by increasing the number of years the explicit use of the term 'bullied' was used and by increasing the variation in type and frequency of bullying that students report experiencing at school. The SCS/NCVS also provides information on parents' economic resources and on school policies and climate. The SCS/NCVS rotational group survey design may lend itself to constructing a short-run longitudinal data set where students are matched across survey years using unique household and person identifiers. Panel data would enable me to account for the role of unobserved student heterogeneity in estimating the correlates of bullying and in estimating the impacts of state laws (the latter requires state-identifiers).[93]

What are the policy implications of this research paper? This study provides preliminary results from thorough analyses of state bullying legislation and highlights the remaining challenges to culminating a conclusive assessment of whether bullying laws work. Empirical research aimed at providing a definitive evaluation of bullying legislation could be facilitated by enhanced data collection, the ability to link data across administrative systems, and broadened data accessibility. Such improvements would enable many other important questions regarding bullying to be investigated, all based on a common and current integrated panel dataset

[93] It is reasonable to assume unobserved student heterogeneity such as a student's endowment of a non-cognitive skill such as resilience is fixed during a short time period like the short panel alluded to in the text.

representative of the K-12 population.[94] A clear and comprehensive understanding of the impacts of various aspects of state bullying legislation can inform U.S. policymakers on the type of legislation that helps ensure all students are protected from bullying at school.

[94] Ideally, one envisions a student-level survey design that includes consistent measures of bullying that disentangle it from other behaviors such as delinquency and also includes a unique household, person, school, and state identifiers. Additionally, one envisions a school-level survey design that includes school policies and reports of other school factors related to climate, crime, academics, etc. and also includes school and location identifiers. Both data sets would be administered in the same academic year and quarter, and each would consist of repeated observations of the same individuals/schools over time. These data could be merged used to construct a panel data set of sufficient sample size by matching via a unique school-identifier.

III SUPPLEMENTAL INSTRUCTION IN DIFFICULT COLLEGE COURSES: A RANDOMIZED CONTROL TRIAL

Introduction

Policymakers now regard college achievement as a practical matter for the masses—not a rite of passage for the privileged few (Leonhardt, 2015; Vara, 2015). In fact, nearly two-thirds of state policymakers, including those in Georgia, have committed to working with their colleges and universities in order to make college completion a top priority.[95] A substantive increase in U.S. college completion rates requires making higher education affordable and accessible to a greater number of students from diverse backgrounds. Policies governing financial aid and student loans may assist with college affordability.[96] And several rigorous studies have examined the impacts of such policies on college enrollment rates [see, for example, van der Klaauw (2002), Stinebrickner & Stinebrickner (2008), and Hoxby & Turner (2015)]. Policy recommendations to address college accessibility often include co-requisite academic support.[97] Yet the impacts of such academic interventions are not well-understood.

Supplemental Instruction (SI) is a voluntary academic support program aimed at improving academic outcomes in historically difficult courses (Blanc et al., 1983).[98] Accessible to all students in targeted courses, SI uses regularly scheduled peer-led review sessions to help students to master content and develop study skills (Blanc et al., 1983). SI has been implemented in at least 230 U.S. post-secondary institutions and in at least 35 colleges and

---

[95] Complete College America is a national nonprofit organization established in 2009 with the intent of helping states increase the number of Americans with college degrees. States participating in this alliance can be found at http://completecollege.org/the-alliance-of-states/.

[96] For example, President Obama recently undertook several initiatives including expanding Pell grants, capping student loan payments based on income, and proposing to make community college free for all (Vara, 2015).

[97] See, for instance, recommendations by Complete College America at http://completecollege.org/about-cca/.

[98] Historically difficult courses typically consist of gateway courses with high rates (in excess of 30%) of withdrawals and D and F grades (Blanc et al., 1983).

universities in Canada, Australia, New Zealand, United Kingdom, Ireland, Sweden, and South Africa combined.[99] Proponents of SI present evidence that SI increases course grades on average, but such estimates are derived from observational designs from which causal inferences are difficult to draw largely because the methods used do not convincingly address selection issues inherent in these data (Blanc et al., 1983; Ramirez, 1997; Loviscek & Cloutier, 1997; Ogden et al., 2003; Lewis et al., 2005; Huynh et al., 2010; Malm et al., 2011; Wilson & Rossig, 2013; Stock et al., 2013).[100]

Despite the widely accepted beliefs about the benefits of SI, few students attend SI review sessions. Such a pattern is not unusual. Most adults express a desire to save more, exercise more, and take up more preventative health care. Notwithstanding these stated preferences, people fail to save, exercise and visit the doctor regularly. Drawing on insights from the behavioral sciences, we devise a simple intervention that aims to encourage greater student use of SI – well-timed and targeted reminders.

We conduct a field experiment that combines practitioner's deep program knowledge, administrative data, and economic research in behavioral interventions and rigorous evaluation designs in order to improve the implementation of the SI program at Georgia State University (GSU).[101] We randomly assign encouragements (i.e., additional email reminders) across a target population of over 17,000 students to boost participation in SI. We then estimate the average causal impacts of SI on academic performance for those induced by the encouragements using

---

[99] Institutions offering SI programs in the U.S. and internationally is accessible at the University of Missouri-Kansas City's International Center for SI website at http://www.umkc.edu/asm/si/programs.shtml.
[100] As of March 2015, we found no relevant studies on peer-tutoring programs on the U.S. Department of Education's What Works Clearinghouse website at http://ies.ed.gov/ncee/wwc/.
[101] The authors collaborated with the Office of Supplemental Instruction at GSU, and obtained Internal Review Board (IRB) approval.

the random encouragements as an instrumental variable for SI participation. Additionally, GSU is a large, urban institution serving a diverse student population with a substantive SI program. Utilizing such a sample markedly facilitates our ability to detect heterogeneous treatment effects and allows us to present causal estimates with strong external validity.

To our knowledge, this is the first large field experiment involving SI and it contributes to the evidence base on multiple dimensions. In addition to shedding light on whether well-targeted reminders can boost participation in academic support programs and whether such participation improves academic performance on average, we also shed light on whether such programs can increase college achievement for important subpopulations.

Rationale

As noted above, estimates of the impact of SI based on observational data likely do not account for selection because students self-select into the SI program as participation is voluntary and all students enrolled in a targeted course can access SI review sessions.[102] Therefore, unobserved student heterogeneity may be driving the estimated positive correlation between SI and academic performance even after accounting for observable demographic and academic characteristics of students. Our research design, however, allows us to estimate the average causal effects of SI—for a subpopulation under certain assumptions—and our sample allows us to detect heterogeneity in such impacts and present estimates with strong external validity.

Nonetheless, numerous colleges and universities around the world have implemented SI in an attempt to "improve student grades in targeted historically difficult courses, increase

---

[102] In future versions of this paper, we will discuss prior studies in more detail and include a table that summarizes previous SI impact studies.

retention with targeted historically difficult courses, and increase graduation rates of students."[103] Indeed, Georgia policymakers have partnered with colleges and universities within the state to "[take] bold actions to significantly increase the number of students successfully completing college. That means achieving degrees and credentials with value in the labor market and closing attainment gaps for traditionally underrepresented populations."[104]

GSU is the second largest university in Georgia (the largest in Atlanta) with roughly 32,000 students, the 14th most diverse university in the nation, its student body is comprised of nearly equal shares of black (38%) and white (37%) students, and it leads the nation in graduating students from widely diverse backgrounds.[105] Additionally, GSU has a large SI program. Roughly 170 SI leaders (i.e., peer tutors) were employed in the Spring and Fall 2014 semesters servicing a potential target population of over 17,000 student combined.[106] Like SI programs at other higher education institutions, GSU's SI program reports that student participation rates are low—around 15%--despite estimates that "students who consistently participate in SI sessions receive, on average, a half or a whole letter grade higher than students enrolled in the same course that do not attend SI sessions" (GSU, 2012).[107]

Why? Are students uncertain about the benefits of SI? Are they forgetful? Do they procrastinate? Recent experiments in health, education, and savings domains reveal similar behavioral paradoxes among adults and show that well-targeted reminders can induce more people to behave in ways that improve their own welfare [see, for example, Milkman et al.

[103] Three purposes of SI were quoted from UMKS SI's webpage at http://www.umkc.edu/asm/si/overview.shtml.
[104] Quote taken from Complete College America's website, http://completecollege.org/the-alliance-of-states/.
[105] Facts taken from GSU's Quick Facts Flyer accessed on 29 March 2015 at http://www.gsu.edu/wp-content/uploads/ 2015/01/QuickFactsFlyer12_14.pdf .
[106] GSU had a total undergraduate enrollment of about 25,000 per semester in Spring and Fall 2014.
[107] These letter grade equivalents are based on difference-in-means estimates of roughly 0.5 course GPA.

(2011) and Lavecchia et al. (2015)]. We combine some of these so-called "nudges" in the context of SI using a simple intervention that could easily be scaled-up.

Experimental Environment and Design

SI Program Offered at GSU

At GSU, the SI program offers free, voluntary academic assistance that utilizes peer-assisted study sessions.[108] SI review sessions are facilitated by peer-tutors, or "SI leaders", who are students themselves; they have previously performed well in the course and serve as model students by attending all class lectures and taking notes (Blanc et al., 1983). SI sessions focus on integrating recent course material with study skills while students work cooperatively with each other (Blanc et al. 1983). No new course content is introduced in SI sessions.

For each course (e.g., Professor Adam Smith's 4 PM section of Econ 2106), SI review sessions are offered outside of class either two (or three) times a week for sixty-five (or forty-five) minutes starting the second week of the semester and continuing until the last day of classes.[109] While student attendance at SI review sessions is recorded, it is important to note that instructors do not have access to SI attendance records.[110] This ensures that students' do not feel coerced to participate in SI or to attend more sessions because they perceive their behavior is being scrutinized by the instructor and thus reflected in their course grade. Throughout the semester, instructors are expected to support and promote SI in their course by

[108] GSU's SI program is modeled in the spirit of the UMKC's Supplemental Instruction program, http://www.umkc.edu/asm/si/index.shtml

[109] In theory, a student enrolled in a course targeted with SI that has regularly scheduled review sessions two times a week for roughly fifteen weeks could attend at most 30 SI review sessions. For classes where SI review sessions are held three times a week, a student could attend at most 45 SI review sessions in a given semester.

[110] At the end of the semester, the Office of SI provides every instructor with a report showing the raw difference-in-means between SI and non-SI students' final course grades separately for each course. More specifically, the instructor receives two of these reports per course; one where SI participation is defined as attending 2 or more SI visits and another where SI participation is defined as attending 5 or more visits.

allowing SI leaders opportunities to make announcements or give reminders about times and locations of SI review sessions in class and/or via university-integrated course websites (GSU, 2012).

Random Encouragement Design

Phase1: Devising Encouragements

We devise an intervention (i.e., encouragements) that would be nearly costless to implement and could easily be scaled up based on insights from behavioral economics coupled with feedback from the Office of SI. The encouragements are personalized and consist of additional email reminders with the day/time/location of the SI review sessions scheduled for that week for a given course, as well as information about the benefits of SI and a prompt for students to commit to attending SI sessions (see sample email reminder message in Appendix Figure C1). The additional email reminders are sent at 8 PM on the eve of each scheduled SI review session from a GSU mass email account named SIbroadcast@gsu.edu. Hence, our treatment incorporates strategies that work through four channels: (i) disseminating information about SI reduces search costs and uncertainty over the benefits of the program, (ii) regular timely email reminders about SI sessions assist with forgetfulness, (iii) the planning prompt serves as a commitment device and helps with procrastination, and (iv) personalizing the invitation and sending it from a high power source may improve response rates.[111] We combined all four because we wanted to maximize the statistical power of our design and because we were not trying to contrast the effects of different components.

[111] Joinson & Reips (2007) study the effect of personalized salutation and sender power on signing up to an online survey panel, and subsequent survey response rates. They find a significant salutation effect when the power of the sender was high, but not when power of the sender was neutral.

Phase 2: Evaluating the Encouragements

The main objective of Phase 2 is to estimate the average causal impact of the encouragements on student participation in SI under certain assumptions (discussed in the Methodology section), where SI participation is measured as the number of SI review sessions attended in a given semester (a count variable).[112]

We randomly assign students enrolled in courses targeted with SI to receive the encouragements (i.e., additional email reminders); these students are referred to as the treated group. Those students that do not receive the encouragements serve as the control group. Because random assignment creates variation in the encouragement that is uncorrelated with the other independent variables (subject area, SI leader, observable student characteristics, etc.) and with the error term (i.e., unobservable student characteristics), the difference in mean SI outcomes between the treatment and control groups is attributed to the effect of the encouragements (or to sampling variability).

Incorporating data on student demographic and academic characteristics improves the evaluations in both Phase 2 and Phase 3 by increasing the statistical power of the analysis (increasing the precision of the impact estimates) and by allowing the Office of SI to better understand variation in student responses to the email reminders. In addition, the results from Phase 2 per se may provide data-driven ways to increase SI participation rates at GSU. Lastly, results from Phase 2 allow us to examine whether there is sufficient power to carry out an evaluation of SI in the Phase 3.

[112] The Office of SI at GSU measures SI participation using two separate binary measures; one SI participation measure equals one if the number of SI review sessions attended is two or more and zero otherwise, while the other measure equals one if the number of SI review sessions attended is five or more and zero otherwise. Alternatively, one could construct a binary measure of SI participation that is equal to one if the number of SI review sessions attended is one or more and zero if no SI review sessions were attended.

Phase 3: Evaluating SI

The main objective of Phase 3 is to estimate the average causal effect of SI on course grade for those students whose participation is affected by the encouragements under certain assumptions, which we discuss in the Methodology section.

In the absence of the random encouragement intervention discussed in Phase 2, SI participation is non-random. SI is a voluntary academic support program so students decide how many, if any, SI review sessions they will attend. It is likely that unobserved student characteristics such as motivation and true ability are correlated with both participation and course grade. Because such factors are difficult to adequately measure (although certain student characteristics may account for a substantive amount of variation in both academic outcomes and SI participation), estimates of the impact of SI derived from observational data alone do not convincingly account for self-selection.

We use randomization to deal with the self-selection problem. As discussed in Phase 2, students are randomly assigned to receive the encouragements. It is reasonable to assume the additional email reminders only affect academic outcomes (i.e., course grade) through the SI participation channel. And the random encouragements are uncorrelated with unobserved factors that may be correlated with SI participation or course grade like motivation or true ability. Thus, we employ a Two-Stage Least Squares (TSLS) estimation strategy that allows us to use all the information available on SI participation—the number of SI review sessions attended, instead of a dummy variable for SI participation—to analyze experimental-participation data merged with additional student characteristics and course grades. The TSLS approach allows us to estimate the average causal response (ACR) of SI under certain assumptions discussed in the Methodology section (Angrist & Imbens, 1995).

Data

Our experiment was conducted in the Fall 2014 semester across a target population of about 10,000.[113] We started sending email reminders on the third week of classes and continued for the duration of the semester for a total of 11 weeks.[114] Randomization was done within each class wherein half of the students were randomly assigned the treatment (reminders) and the remaining students served as the control (no reminders) group. In addition to treatment status, the experimental data consisted of semester, year, student name, university email address, course information and corresponding SI review session information such as SI leader name and the day, time, and location of the weekly scheduled sessions.

We obtained administrative data on SI attendance at the end of the semester. There are 9,866 observations with both treatment status and SI participation data. Because SI review sessions and the email reminders do not commence at the same time and because students may respond differently to the encouragements, we use the SI attendance records to create variables allowing us to account for pre-treatment attendance behavior within a semester (i.e., the number of SI review sessions attended during the first two weeks of SI before the encouragements began). Recall our SI outcome of interest is a count variable measuring the number of SI review sessions. Hence, we utilize two SI count variables in our study, a pre-treatment SI participation measure and a (post-treatment) SI outcome measure.

[113] A similar field experiment was also conducted in Spring 2014 across a target population of about 8,000 students, but caveats exist with including this data in the evaluation; see Extensions section for discussion. The Spring 2014 study was also done in collaboration with Office of SI and IRB approval was obtained.
[114] It took about two weeks to prepare the student data for mass email broadcasts.

We also obtained administrative data on demographic (age, sex, race, ethnicity, citizenship status, and whether a student lives on/off campus) and academic information (course grade, college/institutional GPA, level, Pell eligibility, first generation college goer status, high school attended, and Freshman Index score).[115] We transform students' final course grades from an ordered categorical measure on a plus-minus letter grade scale (A+, A, A-, B+,…, C, C-, D, F) to its numerical equivalent on a 4.3 scale (4.3, 4.0, 3.7, 3.4, …, 2.0, 1.7, 1.0, 0) in order to create a discrete numerical measure of academic performance facilitating both the interpretation and a discussion of our estimates in the context of a broad literature.[116]

The administrative data was matched to experimental data.[117] There are 8,879 out of 9,866 observations in Fall 2014 for which we have grade, experimental and SI participation data.[118] We examine empirically which covariates explain most of the variation in outcomes. Although all explorations are not shown, selected results are presented in Appendix Table C2 for SI participation and Appendix Table C3 for course grades. SI participation in the pre-treatment period explains over 40% of the variation in (post-treatment) SI participation. High school GPA explains about 10% of the variation in course grade. Dummy variables for sex and race/ethnicity explain little variation (an additional 1-2%) in SI participation or course grade, however, adding such covariates improves the models fit (lower Akaike Information Criterion).

[115] College GPA is a student's cumulative institutional GPA at the start of the corresponding semester; classification refers to whether a student is a Freshman/Sophomore/Junior/Senior; Freshman Index score is a GSU-created score created when a student enrolls as a freshman that combines their HS GPA with their SAT or ACT score. Pell eligibility only informs us of whether a student was eligible for a Pell grant award conditional on having applied for financial aid (i.e., completed and submitted FAFSA form); roughly 80% of GSU students apply for financial aid. Additionally, first-generation status is based parents educational attainment as reported in FAFSA data.
[116] Full scale available http://registrar.gsu.edu/academic-records/grading/cumulative-grade-point-average/.
[117] Experimental data and student covariates were merged with the assistance of the Office of SI and the Office of Institutional Research at GSU using a unique student identifier common to both data sets.
[118] The resulting experimental-administrative student-level data set is used in Phases 2 and 3. Only descriptive and empirical results from the analyses are discussed with Professor Ferraro. Since instructors do not have access to any individual-level data linking students to SI attendance or participation, students' anonymity is preserved.

Only 6,171 observations have data on key student covariates (i.e., high school GPA, sex, race/ethnicity).[119] About 30% of the original sample is lost due mostly to missing high school GPA, however, the statistical power increases (precision of impact estimates increases) markedly.[120]

Table 28 presents summary statistics for the sample. Note the unit of observation is at the student-class level since the same student could have been enrolled in multiple classes targeted with SI in a particular semester (2,635 observations, or 43%, are "duplicates"). For ease of discussion, however, we may refer to observations as students. The average number of SI review sessions attended (henceforth, SI visits) per class is 1.8 and approximately 63% of the sample attend no SI review sessions for a given class for the entire semester (i.e., makes 0 SI visits). The average course grade is 2.74, or roughly a "B minus", on the 4.3 letter grade scale at GSU, whereas the average high school GPA is 3.36, or roughly a "B plus", on a 4.0 letter grade scale.[121]

There are more observations in the sample characterized by being a black student (42%) than being a white student (29%). That is, black students are more likely to be enrolled in classes offering SI than are white students. And more observations are characterized by being a female student (59%) than being a male student (41%); female students are more likely than

[119] Summary statistics for the sample before dropping observations due to missing data are presented in Appendix Table C3.
[120] Appendix Tables C4 and C5 present impact estimates with and without the HS GPA covariate. Neither the estimates nor their standard errors change much. In later versions of the paper, we will discuss why this loss occurs (e.g., administration data collection procedures) and provide some descriptive evidence showing how students without HS GPA data may differ in observable ways from students with HS GPA data as well as impact estimates based on the larger sample that does not exclude observations missing data on HS GPA.
[121] We have data on high school attended. The 4.0 grading scale is used in most secondary schools from which students in the sample graduate.

male students to be enrolled in courses with SI. But among the black share of observations in the sample, 67% are female and 33% are male. In other words, black females are twice as likely as black males to be enrolled in classes with SI.

Table 28. Summary Statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Grade	6171	2.7416	1.0681	0	4.3
Treatment	6171	0.4988	0.5000	0	1
SI Participation (number of SI visits)	6171	1.7953	4.3312	0	34
SI Participation (Pre-Treatment)*	6171	0.3230	0.9667	0	8
HS GPA*	6171	3.3641	0.3548	0	4
Male*	6171	0.4090	0.4917	0	1
White	6171	0.2936	0.4555	0	1
Black*	6171	0.4221	0.4939	0	1
Asian*	6171	0.2175	0.4126	0	1
Other*	6171	0.0668	0.2496	0	1
Hispanic*	6171	0.0859	0.2802	0	1
Age (years)	6171	19.8338	2.0784	14.6	60.2
US citizen	6171	0.9152	0.2785	0	1
Non-citizen	6171	0.0848	0.2785	0	1
Freshman	6171	0.4204	0.4937	0	1
Sophomore	6171	0.3748	0.4841	0	1
Junior	6171	0.1330	0.3396	0	1
Senior	6171	0.0619	0.2410	0	1
Other level	6171	0.0099	0.0989	0	1
Institutional GPA	4073	3.1166	0.6050	0	4.3
Freshman Index	1759	2748.5	231.7	2051	3474
Pell-eligible	6171	0.6010	0.4897	0	1
First-generation college goer	4878	0.2612	0.4393	0	1

*indicates a key covariate.

Methodology

First, we estimate the average causal effects of the encouragements on the number of SI visits using a raw difference-in-means between the treated (i.e., those students who are sent additional email reminders) and control (i.e., those students who are not sent additional email reminders) groups. Second, we use the encouragements as an instrumental variable for SI participation in order to estimate the average per-session causal effect of SI on course grade for those students induced by encouragements.

To do so, we use a Two-Stage Least Squares (TSLS) estimation strategy and the regression models provided by equations (1) and (2).[122]

$$SI_i = \alpha_0 + \gamma_0 Z_i + \beta_0 X_i + \upsilon_i \qquad (1)$$

$$grade_i = \alpha_2 + \delta SI_i + \beta_2 X_i + \varepsilon_i \qquad (2)$$

Equation (1) represents the first-stage regression model, SI_i is the observed SI outcome for student *i*. X_i is a vector of demographic and academic characteristics for student *i* and υ_i is an idiosyncratic error term. Z_i is the encouragement (treatment) dummy variable, which equals 1 if student *i* was randomly assigned to be sent the additional email reminders. γ_0 captures the average causal effect of the encouragements on SI participation under certain assumptions discussed below. Equation (2) represents the second-stage regression model. $grade_i$ is the observed course grade for student *i*. ε_i is an idiosyncratic error term. δ captures a weighted (constant) per-session average causal effect of SI on course grade, for those students whose SI participation is affected by the encouragements under certain assumptions

[122] Angrist & Imbens (1995) show how TSLS cab be used to estimate a weighted (uniform) per-unit average causal effect for a subpopulation.

discussed below. Assuming a constant average treatment effect may be a strong assumption in the context of SI. It is likely that the encouragements induce students to respond differently, and discuss this concern later in the paper.[123]

The TSLS estimation approach estimates the first-stage equation, equation (1), using OLS. Then using predicted values of the outcome in the first stage, $\widehat{SI_i}$, as the key covariate in equation (2), the causal equation of interest, equation (2), is estimated using OLS.[124] We also cluster standard errors by student because even though we randomize a certain number/percent of the total number of students within each class, students may be enrolled in multiple classes targeted with SI so outcomes are likely correlated across observations.[125]

When estimating raw difference-in-means, equations (1) and (2) do not include additional covariates so X_i is omitted and the regression model is saturated. Thus, it is appropriate to estimate the model using OLS (Angrist, 2001; Angrist & Pischke, 2009 p.51). It is important, however, to include student covariates since some of the variation in student participation and course grade can be attributed to observable demographic and academic characteristics that we can control for using administrative data. Yet one needs to consider the tradeoff between the amount of variation explained versus how parametric the model becomes when including additional covariates. If indeed there are few covariates and they are discrete,

[123] In the Robustness Check section, we discuss investigating other approaches that may allow us to identify causal estimands other than a uniform per-session average treatment effect of SI. Providing valid causal estimates of such variation in responses could be quite informative to both practitioners and researchers.
[124] We use ivregress 2lsl command in Stata (we do not explicitly estimate each stage) so standard errors and test statistics obtained are valid.
[125] Recall observations are at the student-class level. Students self-select into classes. If students enroll in particular classes in a way that is systematically related to unobservable student characteristics that may influence the SI participation or course grade (e.g., preferences for days/times the class meets), then imposing a group structure on the standard errors (clustering) allows us to obtain reliable standard errors for inference (Angrist & Pischke, 2009).

then a linear regression model is appropriate regardless of the distribution of the SI outcome (Angrist, 2001). Since we utilize one continuous covariate (high school GPA), one discrete covariate (pre-treatment SI participation), and four dummy variables (male, black, Asian, other race), OLS estimation should be appropriate.

To obtain estimates of the average causal effect of the encouragements on the SI outcomes, the following identification assumptions must hold:

(i) Stable Unit Treatment Value Assumption (SUTVA): non-interference across students.

(ii) Independence Assumption: Z is independent of all potential outcomes and potential treatment intensities

These identification assumptions are not testable. We discuss the plausibility of these assumptions below. Assumption (i) means that SI participation for a given student remains stable regardless of which other students happen to be treated (Gerber & Green, 2012). We do not deny that students receiving the encouragement likely interact with students who do not receive the encouragements. One could surmise that a student in the treatment group might decide to forward each additional email reminder to a student in the control group (e.g., a classmate) and, as a result, this control group student attends more SI review sessions. Such spillovers would introduce downward bias in our estimate of the average impact of the encouragements on SI participation, making our estimate a lower bound for the average causal impact of the encouragements. However, we feel it is unlikely that students receiving the encouragements interact with other students (in either the treatment or the control group) in such a way that interferes with or changes their decisions to attend SI review sessions.[126]

Assumption (ii) requires that the encouragements are uncorrelated with the error terms in Equations (1) and (2). Because we randomized the encouragements within each class and because some students may be enrolled in more than one class targeted with SI, some observations are correlated across classes.[127] Such correlation does not bias our estimator of the average causal impact, but it does bias its variance estimator. If we relax Assumption (ii) and allow for within student correlation, then the clustered sandwich estimator is unbiased. Hence, we cluster standard errors by student, assuming the encouragements are uncorrelated with the error terms in Equations (1) and (2) across students—not across observations (student-class).

We compare the means and standard deviations of the key covariates between treatment and control groups. If the two groups are similar in most observable and relevant student characteristics that influence participation (e.g., academic characteristics, pre-encouragement behavior, sex, race/ethnicity), then it could be that unobservable student characteristics (innate ability, motivation, etc.) may be distributed similarly between the treated and control groups as well. Table 29 provides means and standard deviations of key covariates and outcomes in the sample. There are no differences in means of the covariates between the two groups.[128] In fact, the means and standard deviations of other observed

[126] To affect average causal impact of the encouragements, there would need to be sufficient number of instances of diligent email forwarding between treatment and control group students within the same class, which does not seem reasonable. A more plausible threat (discussed later) is that the encouragement may not being “strong enough” to induce a sufficient number of students to go to more SI review sessions.
[127] The unit of observation is student-class, so there is correlation across classes
[128] Difference in means were tested using non-parametric (Mann-Whitney) test.

Table 29. Sample Means, by Treatment Status. N = 6,171

	Control Group, N=3,093	Treatment Group, N=3,078
Treatment	0 (0.0)	1 (0.0)
Grade	2.7369 (1.0776)	2.7464 (1.0587)
Number of SI visits	1.6360 (4.1445)	***1.9555 (4.5060)
Number of SI visits, pre-Treatment	0.3107 (0.9321)	0.3353 (1.0002)
HS GPA	3.3653 (0.3550)	3.3628 (0.3546)
Male	0.4087 (0.4917)	0.4094 (0.4918)
White	0.2907 (0.4541)	0.2966 (0.4568)
Black	0.4242 (0.4943)	0.4201 (0.4937)
Asian	0.2156 (0.4113)	0.2193 (0.4138)
Other	0.0695 (0.2544)	0.0640 (0.2448)
Hispanic	0.0860 (0.2804)	0.0858 (0.2801)

***indicates means are different at 1% level of significance based on non-parametric (Mann-Whitney) test.

Table 30. Falsification Test: Effect of the Encouragements on SI Participation (Pre-Treatment)

Dependent Variable:	SI Participation (Pre-Treatment)	
Covariates	(1)	(2)
=1 if in treatment group	0.0246	0.0254
	(0.0244)	(0.0244)
High School GPA		0.1812***
		(0.0373)
Male		-0.1188***
		(0.0257)
Black		0.0525*
		(0.0301)
Asian		0.0836**
		(0.0377)
Other race		0.0655
		(0.0564)
Hispanic		-0.0902**
		(0.0396)
Observations	6,171	6,171
Adj. R-squared	-3.90e-07	0.0109

Robust standard errors in parentheses, clustered by student id. *** p<0.01, ** p<0.05, * p<0.10

student and academic characteristics (not shown) are also similar between the treatment and control groups. Hence, the randomization appears to have worked based on observables.

One might argue, however, that covariate balance is inappropriate, particularly in the case of selection-on-unobservables. Therefore, we perform a falsification test, or test of known effect. We use our pre-treatment SI participation variable as the outcome of interest in order to test the null hypothesis that the random encouragements are uncorrelated with SI participation prior to the implementation. Table 3 shows the regression results from estimating Equation (1) where the pre-treatment number of SI visits is the dependent variable. The encouragements (treatment group status = 1) have no estimated effect on SI participation without (column 1) or with (column 2) controls for academic and student characteristics.

To obtain estimates of a weighted (constant) per-unit average causal effect of SI participation on course grade for those induced by the encouragements, an additional identification assumption must hold:

(iii) Monotonicity: $(SI|Z = 1) - (SI|Z = 0) \geq 0 \; or \; (SI|Z = 1) - (SI|Z = 0) \leq 0$ for all

In the context of SI, the Monotonicity Assumption means that the randomized encouragements never decrease SI participation. That is, students who are sent the additional email reminders attend at least as many SI review sessions for a class as they would have attended had they not been sent the reminders for the class. This assumption seems reasonable in random encouragement designs (Angrist & Imbens, 1995).[129]

[129] Nonetheless, in future versions of this paper we will show that the cumulative distribution function (CDF) of SI participation given Z=1 and the CDF of SI participation given Z=0 should not cross; this is a testable implication of the Monotonicity Assumption since SI participation (number of SI visits) is a multivalued treatment (Angrist & Imbens, 1995).

Results

Estimates of the Average Impact of the Encouragements

We begin by examining the effect of the encouragements on SI participation. The top panel of Table 31 displays the first-stage regression results, which estimates the average causal impact of the email reminders on the number of SI visits. Columns 1 and 2 show the estimated impact without and with clustering by student, respectively, and without any covariates; clustering has little effect on the standard errors. Column 3 presents estimates of the causal impact of the encouragements controlling for both the pre-treatment number of SI visits and high school GPA. Column 4 also includes controls for sex and race/ethnicity, where white and female are the omitted categories. Appendix Table C4 provides estimates by alternating the order in which the covariates are introduced.

Email reminders increase the number of SI visits by 0.25 or 15% on average, all else equal. If the reminders were scaled-up to all 6,171 students in the sample, we estimate that collectively students would make 926 more visits to SI review sessions. Hence, the email reminders are effective on average by a practically and statistically (1% level) significant amount.

Estimates of the Average Impact of SI

Now that we have shown the encouragements work and they create a significant amount of exogenous variation in SI participation on average, we examine the effect of SI on academic achievement (course grades) using the encouragements as an IV for SI participation. The bottom panel of Table 31 displays the second-stage regression results, which estimates a weighted uniform per-unit average causal impact of SI participation on grades, for those

Table 31. Estimated Per-Session Average Causal Effect of SI Participation on Grade

First-stage				
Dependent variable:	SI Participation			
Covariates	(1)	(2)	(3)	(4)
=1 if in treatment group	0.3195***	0.3195***	0.2473***	0.2487***
	(0.1102)	(0.1102)	(0.0826)	(0.0824)
SI Participation (Pre-Treatment)			2.9571***	2.9499***
			(0.0942)	(0.0938)
High School GPA			0.2072*	0.1872
			(0.1246)	(0.1262)
Male				-0.0882
				(0.0848)
Black				0.5415***
				(0.1003)
Asian				0.3871***
				(0.1205)
Other race				0.0334
				(0.1556)
Hispanic				0.4145***
				(0.1568)
Observations	6171	6171	6171	6171
Adj. R-squared	0.0012	0.0012	0.4386	0.4413
Second-stage				
Dependent variable:	Course Grade			
Covariates	(1)	(2)	(3)	(4)
SI Participation	0.0295	0.0295	0.0428	0.0347
	(0.0847)	(0.0844)	(0.1037)	(0.1018)
SI Participation (Pre-Treatment)			-0.0792	-0.0473
			(0.3071)	(0.3008)
High School GPA			0.8968***	0.9147***
			(0.0463)	(0.0458)
Male				0.1030***
				(0.0314)
Black				-0.3606***
				(0.0668)
Asian				-0.1709***
				(0.0585)
Other race				-0.1641***
				(0.0566)
Hispanic				-0.0540
				(0.0680)
Observations	6171	6171	6171	6171
Adj. R-squared	0.0096	0.0096	0.0950	0.1200

Robust standard errors in parentheses, clustered by student id in Columns 2-4. *** p<0.01, ** p<0.05, * p<0.10.

students whose SI participation is affected by the encouragements. Columns 1 and 2 show the estimated impact without and with clustering by student, respectively, and without any covariates; clustering has little effect on the standard errors. Column 3 presents estimates of the causal impact of the encouragements controlling for both the pre-treatment number of SI visits and high school GPA. Column 4 also includes controls for sex and race/ethnicity, where white and female are the omitted categories. Appendix Table C4 provides estimates by alternating the order in which the covariates are introduced.

The estimated average per-session causal impact of SI on grades is 0.0347, for those students who are induced by the reminders to increase their SI participation. This translates into a 0.03 standard deviation increase in course grade, however, this causal impact is neither statistically significant nor is it practically significant. Interventions deemed successful in the behavioral economics of education literature find impacts on student grades ranging from 0.12 to 0.65 standard deviations, see Lavecchia et al. (2015) for some examples.[130]

There are some caveats for comparing our causal impacts to estimates in previous SI studies. Unlike many previous impact studies of SI often, we use a count variable to measure SI participation by the number of review sessions attended—not a binary measure indicating whether or not a student attended a specific number of SI review sessions. Second, our causal estimates are based on a subpopulation of students in classes targeted with SI—those students who are induced by the encouragements to increase their SI participation. Third, we assume that all students affected by the encouragements respond in the same way by increasing their

[130] Below we discuss some ways in which we will try to compare estimates from our analyses to findings from previous SI impact studies.

SI participation uniformly (hence the per-session average causal effect of SI), which is admittedly a strong assumption ensuing a further discussion below.

With this in mind, we wanted to put our results in perspective for the reader. It is informative to compare our estimates to average marginal "effect" of SI participation on course grade in the absence of the encouragement. So we used OLS to estimate Equation (2) using only observations in the control group. These results are presented in Table 5. We estimate that attended an additional SI review session increases course grade by 0.0263, or 0.02 standard deviations. The estimated positive correlation between SI participation and grade is highly statistically significant, consonant (at least in sign and statistical significance) with findings based on observational research designs.[131]

Robustness Checks

Treatment Effect Heterogeneity

We are also interested in the effects of the encouragements and the marginal effects of an additional SI review session on important subgroups of students (first generation college goers, low income, minorities, etc.). Estimation of treatment effect heterogeneity plays a crucial role in uncovering subpopulations that benefit from (or are unaffected by) treatment; designing optimal treatment plans; testing for the existence or lack of heterogeneous treatment effects; and making generalizations from an experimental sample to a target population (Imai & Ratkovic, 2013). Therefore, we plan to estimate conditional average causal impacts of the encouragements and of SI participation in subsequent analyses.

---

[131] In Robustness Check section, we discuss other ways of possibly providing comparisons to estimates found in previous work on SI.

Table 5. Estimated Correlation Between SI Participation and Grade, Control Group

Dependent variable:	Course Grade			
Covariates	(1)	(2)	(3)	(4)
SI Participation	0.0263***	0.0263***	0.0221***	0.0263***
	(0.0047)	(0.0041)	(0.0055)	(0.0054)
SI Participation (Pre-Treatment)			0.0035	0.0008
			(0.0258)	(0.0252)
High School GPA			0.9127***	0.9168***
			(0.0557)	(0.0567)
Male				0.0758*
				(0.0407)
Black				-0.4326***
				(0.0479)
Asian				-0.2055***
				(0.0563)
Other race				-0.1363*
				(0.0729)
Hispanic				-0.0460
				(0.0699)
Constant	2.6939***	2.6939***	-0.3721*	-0.1815
	(0.0207)	(0.0222)	(0.1917)	(0.2031)
Observations	3,093	3,093	3,093	3,093
R-squared	0.0102	0.0102	0.1005	0.1310
Adj. R-squared	0.0099	0.0099	0.0996	0.1290

Standard errors in parentheses, clustered by student id in Columns 2-4. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Alternative Estimation Strategies

In our analysis, we assume a constant per-unit SI effect, for the subgroup of students induced by the encouragements to increase their SI participation. The linear-in-SI assumption is rather restrictive and possibly is unrealistic. The average impact of attending an additional SI review session is likely to have different impacts on course grade depending on the number of SI review sessions a student would have attended in the absence of the encouragement. For instance, the average causal effect on course grade for students induced by the encouragements to attend one instead of zero SI review sessions may be substantively different from the average causal effect on course grade for students induced to attend eleven instead of ten SI review sessions. Therefore, we are investigating approaches that allow us to estimate (variable) per-session average treatment effects.[132]

Comparisons to Previous Evidence

Many previous impact studies use a binary measure of SI participation, where participation equals one if a student attends one or more SI review sessions and zero otherwise. Some studies also analyze alternative binary measures of SI participation where participation is defined as two or more SI visits or five or more SI visits. We also plan to perform our analysis separately using each of these binary measures of SI participation. Our analysis utilizes the number of SI visits specifying SI participation as a multi-valued treatment, whereas other studies often misspecify SI participation as a binary treatment. Therefore, it would be informative to show to what extent incorrectly parametrizing SI participation results in a

[132] A potentially promising paper for dealing with variable per-unit average treatment effect is Lochner & Moretti (2015) given that we have a binary IV and multi-valued treatment.

relatively larger estimate of the effect of SI, for the subpopulation induced by the encouragements, as compared to the average per-unit effect.[133]

Conclusion

Supplemental Instruction (SI) is a voluntary, peer-led academic assistance program aimed at increasing academic achievement in difficult college courses. Proponents of SI base claims of the program's success on anecdotes and evidence from observational studies. While SI is widely used, it has been poorly evaluated. Our research design overcomes problems arising from students' self-selection into SI. We conduct a field experiment at a large urban university having a large SI program and serving a diverse student population (Georgia State University). Drawing on insights from behavioral economics, we devise encouragements to increase student participation in SI. Next we randomly select students to receive the additional email reminders to attend SI. We then use these random encouragements as an instrumental variable for SI participation in order to estimate a uniform per-session average causal impact of SI on course grades, for those students induced by the encouragements.

First, we find that additional email reminders are effective at increasing student participation in SI by 15% on average, controlling for academic and demographic student characteristics. Second, we find a positive and highly statistically significant relationship between SI and grade when we employ an observational research design, yet we find no (statistically or practically) significant per-session average causal impact of SI participation on course grade among students responding to the encouragements. Caveats in our approach are acknowledged.

---

[133] See, for example, Angrist & Imbens (1995) for further discussion of incorrectly specifying variable treatment intensity as a binary treatment.

In subsequent analyses, we investigate alternative identification strategies that allow us to identify the per-session average causal impact of SI on grades when these average marginal effects vary by the number of SI review sessions attended, weakening the constant per-session assumption. Additionally, the large, diverse setting facilitates investigating heterogeneity in treatment effects across important subgroups (first-time college goers, low income, minorities, etc.).

Our study is the first large-scale field experiment involving SI, enabling us to present convincing causal evidence on SI. It are also the first study to find a data-driven way to significantly increase student participation in SI, devising an encouragement that is nearly costless to implement and can easily be scaled-up. Our results concerning the marginal average causal impact of SI participation on course grade are compelling. They highlight the importance of adequately dealing with self-selection, which may likely be driven by unobservable student characteristics, and may provide a cautionary example for practitioners and researchers evaluating academic support programs. Lastly, our sample allows us to provide causal estimates with strong internal and external validity. These findings are informative for the Office of SI at GSU and the larger academic community because SI is widely used and co-requisite academic support is a predominant policy recommendation for increasing the number of Americans with college degrees—a timely national priority.

Extensions

Duration of the Encouragements

Our experiment was initially conducted in Spring 2014. Unlike Fall 2014, however, the additional email reminders were sent only for the last six weeks of the semester. The late start (about mid-semester) in Spring 2014 largely due to an unanticipated delay in IRB review process. So we and our collaborators agreed to (in some sense) replicate the experiment in Fall 2014 for the full semester. Altogether we have over 17,000 observations from Spring and Fall 2014. Arguably Spring and Fall semesters may differ in other important ways in general (student effort, learning by doing, updated beliefs about SI benefits, etc.). Nonetheless, we could extend our analysis to compare estimated average treatment effects between Spring and Fall. Examining whether the duration of the encouragements matters could provide additional information to the Office of SI and aid in their decision making as to whether, and to what extent, scaling up the encouragements could improve how SI is implemented at GSU.

A APPENDIX TO CHAPTER I

Table A1. Coefficients (s.e.) for O*NET D/T Attributes and Selected Control Variables to Accompany Table 5 Regression Results for Women and Men

	Women			
Independent Variables	(1)	(2)	(3)	(4)
Developing and Building Teams				-0.0185
				(0.0205)
Training and Teaching Others				-0.0745***
				(0.0204)
Coaching and Developing Others				0.0164
				(0.0250)
Instructing				-0.0466***
				(0.0169)
Married, spouse present	0.0407***	0.0416***	0.0422***	0.0428***
	(0.0054)	(0.0048)	(0.0048)	(0.0044)
Separated/widowed/divorced	0.0116***	0.0099**	0.0107***	0.0111***
	(0.0041)	(0.0041)	(0.0039)	(0.0039)
Black	-0.0718***	-0.0752***	-0.0741***	-0.0727***
	(0.0095)	(0.0090)	(0.0093)	(0.0083)
Other race	-0.0155*	-0.0189**	-0.0192***	-0.0217***
	(0.0082)	(0.0073)	(0.0072)	(0.0066)
Hispanic	-0.0803***	-0.0788***	-0.0779***	-0.0767***
	(0.0070)	(0.0069)	(0.0071)	(0.0068)
Foreign-born citizen	-0.0367***	-0.0400***	-0.0399***	-0.0394***
	(0.0079)	(0.0077)	(0.0077)	(0.0075)
Foreign-born noncitizen	-0.1313***	-0.1301***	-0.1316***	-0.1287***
	(0.0125)	(0.0125)	(0.0127)	(0.0126)
Experience	0.0381***	0.0384***	0.0382***	0.0377***
	(0.0034)	(0.0034)	(0.0034)	(0.0034)
Experience^2	-0.0017***	-0.0017***	-0.0017***	-0.0017***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Part-time (hours<35)	-0.0966***	-0.1005***	-0.0975***	-0.0961***
	(0.0150)	(0.0166)	(0.0150)	(0.0144)
Public	0.0363	0.0743***	0.0672***	0.0901***
	(0.0296)	(0.0245)	(0.0254)	(0.0268)
Private not-for-profit	-0.0207	-0.0112	-0.0129	-0.0157
	(0.0221)	(0.0202)	(0.0213)	(0.0200)
Union member	0.1185***	0.1283***	0.1301***	0.1376***
	(0.0100)	(0.0103)	(0.0100)	(0.0113)

Standard errors in parentheses, clustered by occupation. *** p<0.01, ** p<0.05, * p<0.10. See Table 5 notes.

(Table A1 continued on next page)

Table A1 (continued).Coefficients (s.e.) for O*NET D/T Attributes and Selected Control Variables to Accompany Table 5 Regression Results Women and Men

	Men			
Independent Variables	(1)	(2)	(3)	(4)
Developing and Building Teams				-0.0069
				(0.0120)
Training and Teaching Others				-0.0183
				(0.0190)
Coaching and Developing Others				-0.0172
				(0.0192)
Instructing				-0.0184
				(0.0139)
Married,spouse present	0.1249***	0.1252***	0.1244***	0.1241***
	(0.0051)	(0.0054)	(0.0052)	(0.0052)
Separated/widowed/divorced	0.0416***	0.0405***	0.0409***	0.0409***
	(0.0054)	(0.0055)	(0.0054)	(0.0054)
Black	-0.1319***	-0.1370***	-0.1311***	-0.1312***
	(0.0074)	(0.0077)	(0.0075)	(0.0075)
Other race	-0.0421***	-0.0425***	-0.0433***	-0.0437***
	(0.0104)	(0.0106)	(0.0100)	(0.0097)
Hispanic	-0.1149***	-0.1172***	-0.1136***	-0.1132***
	(0.0075)	(0.0077)	(0.0075)	(0.0075)
Foreign-born citizen	-0.0464***	-0.0465***	-0.0471***	-0.0475***
	(0.0087)	(0.0083)	(0.0084)	(0.0083)
Foreign-born noncitizen	-0.1325***	-0.1292***	-0.1317***	-0.1316***
	(0.0112)	(0.0115)	(0.0109)	(0.0109)
Experience	0.0350***	0.0349***	0.0350***	0.0351***
	(0.0021)	(0.0022)	(0.0022)	(0.0021)
Experience^2	-0.0011***	-0.0011***	-0.0011***	-0.0011***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Part-time (hours<35)	-0.1588***	-0.1647***	-0.1569***	-0.1562***
	(0.0147)	(0.0154)	(0.0149)	(0.0151)
Public	0.0937**	0.0945*	0.0967**	0.1023**
	(0.0468)	(0.0487)	(0.0437)	(0.0440)
Private not-for-profit	-0.1023***	-0.1072***	-0.0961***	-0.0989***
	(0.0352)	(0.0372)	(0.0324)	(0.0331)
Union member	0.1886***	0.1839***	0.1907***	0.1926***
	(0.0136)	(0.0142)	(0.0131)	(0.0127)

Standard errors in parentheses, clustered by occupation. *** p<0.01, ** p<0.05, * p<0.10. See Table 5 notes.

Table A2. Wage Level Regression Estimates for Care Work Effects Using the Longitudinal Sample of Ind/Occ Switchers, Initial Year of 2003/4 – 2008/09 Panels

	Women			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	-0.0604***		-0.0523***	
	(0.0126)		(0.0126)	
Develop/Teach Index		-0.0507***	-0.0362***	
		(0.0121)	(0.0114)	
Assisting and Caring for Others				-0.0102
				(0.0136)
Concern for Others				-0.0330***
				(0.0108)
Job Skills Index	0.1580***	0.1843***	0.1872***	0.1889***
	(0.0112)	(0.0170)	(0.0159)	(0.0187)
Working Conditions Index	-0.0131	-0.0132	-0.0044	-0.0035
	(0.0173)	(0.0165)	(0.0174)	(0.0178)
CPS controls	Y	Y	Y	Y
Observations	18,981	18,981	18,981	18,981
R-squared	0.4215	0.4194	0.4229	0.4236
	Men			
O*NET Variables/Indices	(1)	(2)	(3)	(4)
Caring Index	-0.0661***		-0.0621***	
	(0.0090)		(0.0088)	
Develop/Teach Index		-0.0303***	-0.0112	
		(0.0092)	(0.0073)	
Assisting and Caring for Others				-0.0329***
				(0.0083)
Concern for Others				-0.0203***
				(0.0072)
Job Skills Index	0.1303***	0.1357***	0.1382***	0.1408***
	(0.0086)	(0.0110)	(0.0100)	(0.0112)
Working Conditions Index	-0.0140	-0.0118	-0.0121	-0.0129
	(0.0100)	(0.0103)	(0.0100)	(0.0098)
CPS controls	Y	Y	Y	Y
Observations	21,689	21,689	21,689	21,689
R-squared	0.4780	0.4750	0.4782	0.4783

Standard errors in parentheses, clustered by occupation. *** p<0.01, ** p<0.05, * p<0.10. Dependent variable is ln(hourly earnings). See Table 5 notes for CPS controls; see text for a detailed description of variables.

Table A3. Coefficients (s.e.) for O*NET D/T Attributes and Selected Control Variables to Accompany Table 7 Wage Change Regression Results—Ind/Occ Switchers

	Women			
Δ Independent Variables	(1)	(2)	(3)	(4)
Δ Developing & building teams				-0.0034
				(0.0050)
Δ Training & teaching others				-0.0136**
				(0.0054)
Δ Coaching & developing others				0.0147**
				(0.0057)
Δ Instructing				0.0030
				(0.0047)
Δ experience^2	-0.0031***	-0.0031***	-0.0031***	-0.0031**
	(0.0012)	(0.0012)	(0.0012)	(0.0012)
Δ part-time (hours<35)	-0.0078	-0.0081	-0.0078	-0.0076
	(0.0075)	(0.0075)	(0.0075)	(0.0075)
Δ union	0.0740***	0.0734***	0.0740***	0.0737***
	(0.0140)	(0.0140)	(0.0140)	(0.0140)
Δ public sector	-0.0012	0.0007	-0.0015	-0.0020
	(0.0164)	(0.0164)	(0.0164)	(0.0164)
Δ private not-for-profit	-0.0181	-0.0180	-0.0182	-0.0186*
	(0.0111)	(0.0111)	(0.0111)	(0.0111)
	Men			
Δ Developing & building teams				0.0060
				(0.0040)
Δ Training & teaching others				-0.0068
				(0.0050)
Δ Coaching & developing others				0.0094*
				(0.0053)
Δ Instructing				0.0041
				(0.0043)
Δ experience^2	-0.0024**	-0.0024**	-0.0024**	-0.0024*
	(0.0012)	(0.0012)	(0.0012)	(0.0012)
Δ part-time (hours<35)	-0.0263***	-0.0262***	-0.0263***	-0.0263***
	(0.0083)	(0.0083)	(0.0083)	(0.0083)
Δ union	0.0835***	0.0826***	0.0830***	0.0831***
	(0.0123)	(0.0123)	(0.0123)	(0.0123)
Δ public sector	0.0118	0.0077	0.0108	0.0118
	(0.0176)	(0.0176)	(0.0176)	(0.0176)
Δ private not-for-profit	-0.0234	-0.0279	-0.0238	-0.0235
	(0.0174)	(0.0174)	(0.0174)	(0.0174)

*** $p<0.01$, ** $p<0.05$, * $p<0.10$. See detailed notes for Table 7. Results for Δexperience^3, Δexperience^4, broad industry and occupation changes, and year-pairs are not shown.

B APPENDIX TO CHAPTER II

Appendix Table B1. State participation in national YRBSS, sample years 1993-2011

state	1993	1995	1997	1999	2001	2003	2005	2007	2009	2011
Alabama	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Alaska										
Arizona	Y		Y	Y	Y	Y		Y	Y	Y
Arkansas	Y	Y	Y			Y		Y	Y	
California	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Colorado	Y	Y	Y		Y				Y	Y
Connecticut		Y	Y				Y			
Delaware		Y				Y				
Florida		Y	Y	Y	Y	Y	Y	Y	Y	Y
Georgia	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Hawaii				Y					Y	
Idaho					Y		Y			Y
Illinois	Y	Y		Y	Y	Y	Y	Y	Y	Y
Indiana					Y	Y	Y	Y		Y
Iowa		Y	Y				Y	Y		
Kansas	Y		Y			Y	Y		Y	Y
Kentucky		Y					Y	Y		Y
Louisiana		Y	Y	Y		Y	Y		Y	
Maine	Y	Y	Y	Y	Y	Y				
Maryland	Y		Y			Y				
Massachusetts	Y	Y	Y		Y	Y	Y	Y		Y
Michigan	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Minnesota	Y						Y		Y	
Mississippi	Y	Y	Y	Y	Y			Y		Y
Missouri	Y	Y		Y	Y	Y	Y	Y	Y	Y
Montana					Y					
Nebraska	Y									
Nevada					Y					
New Hampshire										
New Jersey			Y	Y	Y	Y	Y	Y	Y	Y
New Mexico	Y		Y		Y	Y		Y	Y	
New York	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
North Carolina	Y	Y	Y	Y	Y		Y	Y		Y
North Dakota										
Ohio	Y	Y	Y	Y	Y	Y	Y			
Oklahoma			Y		Y		Y	Y		
Oregon	Y				Y		Y		Y	

(continued on next page)

Appendix Table B1 (continued)

state	1993	1995	1997	1999	2001	2003	2005	2007	2009	2011
Pennsylvania	Y	Y	Y	Y		Y	Y	Y	Y	Y
Rhode Island				Y						
South Carolina	Y		Y	Y		Y	Y			
South Dakota						Y				
Tennessee	Y		Y	Y	Y		Y	Y		Y
Texas	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Utah						Y	Y	Y		
Vermont						Y				
Virginia		Y		Y		Y	Y	Y	Y	Y
Washington	Y	Y	Y		Y		Y		Y	Y
West Virginia	Y				Y		Y	Y	Y	Y
Wisconsin			Y	Y	Y	Y	Y	Y	Y	Y
Wyoming										

“Y” indicates a state was included for the given year.

Appendix Table B2. YRBSS: Unintentional Injuries and Violence, selected item analysis, 1993-2011

Question Refers to sometime during the 12 mo. (or 30 days) before the survey	1993	1995	1997	1999	2001	2003	2005	2007	2009	2011
Did not go to school because they felt unsafe at school or on their way to or from school on at least 1 day (30 days)	X	X	X	X	X	X	X	X	X	X
In a physical fight on school property one or more times	X	X	X	X	X	X	X	X	X	X
Threatened or injured with a weapon on school property one or more times (e.g., a gun, knife, or club)	X	X	X	X	X	X	X	X	X	X
Carried a weapon on school property on at least 1 day (e.g., a gun, knife, or club) (30 days)	X	X	X	X	X	X	X	X	X	X
Had property stolen or deliberately damaged on school property	X	X	X			X	X	X		
The next (2) question(s) ask about bullying. Bullying is when 1 or more students tease, threaten, spread rumors about, hit, shove, or hurt another student over and over again. It is not bullying when 2 students of about the same strength or power argue or fight or tease each other in a friendly way.										
Bullied on school property									X	X
Ever been electronically bullied (including through e-mail, chat rooms, instant messaging, Web sites, or texting)										X

"X" indicates that the question was included in the survey in that year.

Appendix Table B3. Evolution of Bullying and Victimization questions on SCS/NCVS, 1989-2011

year	Question wording
1989	The following questions are about crimes that may have happened to you at school. By “at school”, we mean in the school building, on school grounds, or on a school bus. Be sure to include crimes you have told me about earlier in this interview. During the past six months did anyone take money or things directly from you by force, weapons, or threats at school? How many times did this happen in the last six months? During the past six months, did anyone steal something from your desk, locker, or some other place at school (other than any incidents just mentioned)? How many times did this happen? During the past six months, did anyone physically attack you at school? How many times did this happen?
1995	Same as 1989
1999	In the first part of the interview, we asked you about crimes that happened in the last six months, whether in school or not in school. Here, the focus is on crimes that happened to you at school. By “at school”, we mean in the school building, on school grounds, or on a school bus. Did anyone attack, threaten to attack, or take something directly from you by force or threats, or steal something from your desk or locker at school in the last six months, that is, since ____ 1st? What happened? Did someone— 1. Attack you? 2. Threaten to attack you? 3. Take something directly from you by force or threats? 4. Steal something from your desk or locker at school? During the last six months, have you been **bullied** at school? That is, have any other students picked on you a lot or tried to make you do things you didn't want to do like give them money? (You may include incidents you reported before.)
2001	During the last six months, have you been **bullied** at school? That is, have any other students picked on you a lot or tried to make you do things you didn't want to do like give them money? (You may include incidents you reported before.) During the last six months, how often has this happened? During the last six months, have you often felt rejected by other students at school? For example, have you felt rejected because other students have made fun of you, called you names, or excluded you from activities? During the last six months, how often has this happened?
2003	Same as 2001
2005	Now I have some questions about what students do at school that make you feel bad or are hurtful to you. We often refer to this as being **bullied**. You may include events you told me about already. During the last six months, has any student bullied you? That is, has another student… a. Made fun of you, called you names, or insulted you in a hurtful way? b. Spread rumors about you or tried to make others dislike you? c. Threatened you with harm? d. Pushed you, shoved you, tripped you, or spit on you? e. Tried to make you do things you did not want to do, for example, give them money or other things? f. Excluded you from activities? g. Destroyed your property on purpose? During the last six months, how often did (this/these things) happen to you?
2007	Same as 2005, except asks “During this school year” instead of “during last six months”
2009	Same as 2007
2011	Same as 2009

Source: Author created based on questionnaire in SCS/NCVS Codebooks.

Appendix Table B4. Sample composition, by Grade Level, SSOCS 2004-2010

	Number of schools[a]	unweighted	weighted
Elementary	2,670	25.3%	59.3%
Middle	3,660	34.8%	18.4%
High	3,750	35.6%	14.2%
Combined:	440	4.2%	8.1%
PreK/K – 12	190	42.3%	47.4%
7 - 12	140	31.5%	30.5%

[a] Unweighted sample entities are rounded to the nearest ten per IES publication policy.
SSOCS sampling weights are used where indicated. Elementary refers to grades K-5; middle refers to grades 6-8; high school refers to grades 9-12; and combined refers to schools with combined grade levels.

Appendix Table B5. Conditional Means of Students' Bullying Experiences at School, SCS/NCVS

	High School			
	SCS/NCVS 2005-2011		SCS/NCVS 1999-2003	
Outcome	Conditional on Bullied at school=1	Conditional on Bullied at school=0	Conditional on Bullied at school=1	Conditional on Bullied at school=0
Absent because felt unsafe	0.021 (0.143)	0.002 (0.043)	0.074 (0.262)	0.005 (0.073)
Fight at school	0.103 (0.304)	0.020 (0.139)	0.161 (0.368)	0.032 (0.176)
Threatened at school	0.184 (0.388)	0.000 (0.139)	--	--
Carried weapon at school	0.037 (0.190)	0.020 (0.136)	0.051 (0.221)	0.017 (0.129)
Electronically bullied	0.203 (0.403)	0.019 (0.)	--	--
Bullied at school	1.0 (0.0)	0.0 (0.0)	1.0 (0.0)	0.0 (0.0)
	Middle School			
	SCS/NCVS 2005-2011		SCS/NCVS 1999-2003	
Outcome	Conditional on Bullied at school=1	Conditional on Bullied at school=0	Conditional on Bullied at school=1	Conditional on Bullied at school=0
Absent because felt unsafe	0.021 (0.144)	0.001 (0.034)	0.051 (0.220)	0.005 (0.072)
Fight at school	0.173 (0.379)	0.031 (0.173)	0.221 (0.415)	0.061 (0.239)
Threatened at school	0.185 (0.388)	0.0 (0.0)	--	--
Carried weapon at school	0.036 (0.187)	0.012 (0.107)	0.024 (0.153)	0.008 (0.088)
Electronically bullied	0.144 (0.341)	0.009 (0.092)	--	--
Bullied at school	1.0 (0.0)	0.0 (0.0)	1.0 (0.0)	0.0 (0.0)

Sample weights provided in the SCS/NCVS are used to calculate the means (standard deviations).

Appendix Table B6. Pairwise Correlations of Bullying Outcomes, YRBSS 1993-2011

	Bullied	ebullied	absent	fight	threat	weapon	Bullied index
Bullied at school	1						
Electronically bullied	0.414	1					
Absent b/c felt unsafe	0.145	0.153	1				
In Fight at school	0.111	0.091	0.129	1			
Threatened at school	0.192	0.172	0.233	0.219	1		
Carried weapon at school	0.066	0.057	0.117	0.211	0.202	1	
Bullied index	0.136	0.116	0.332	0.363	0.461	0.655	1

YRBSS sample weights are used.

Appendix Table B7. Pairwise Correlations of Bullying Outcomes, High school, SCS/NCVS 2005-2011

	Bullied	ebullied	absent	fight	threat	weapon
Bullied at school	1					
Electronically bullied	0.324	1				
Absent b/c felt unsafe	0.157	0.096	1			
In Fight at school	0.143	0.117	0.042	1		
Threatened at school	0.379	0.271	0.187	0.208	1	
Carried weapon at school	0.054	0.069	0.049	0.095	0.081	1

SCS/NCVS sample weights are used. High school refers to grades 9-12.

Appendix Table B8. Pairwise Correlations of Bullying Outcomes, Middle school, SCS/NCVS 2005-2011

	Bullied	ebullied	absent	fight	threat	weapon
Bullied at school	1					
Electronically bullied	0.282	1				
Absent b/c felt unsafe	0.143	0.123	1			
In Fight at school	0.190	0.102	0.083	1		
Threatened at school	0.360	0.250	0.163	0.220	1	
Carried weapon at school	0.053	0.069	0.004	0.065	0.089	1

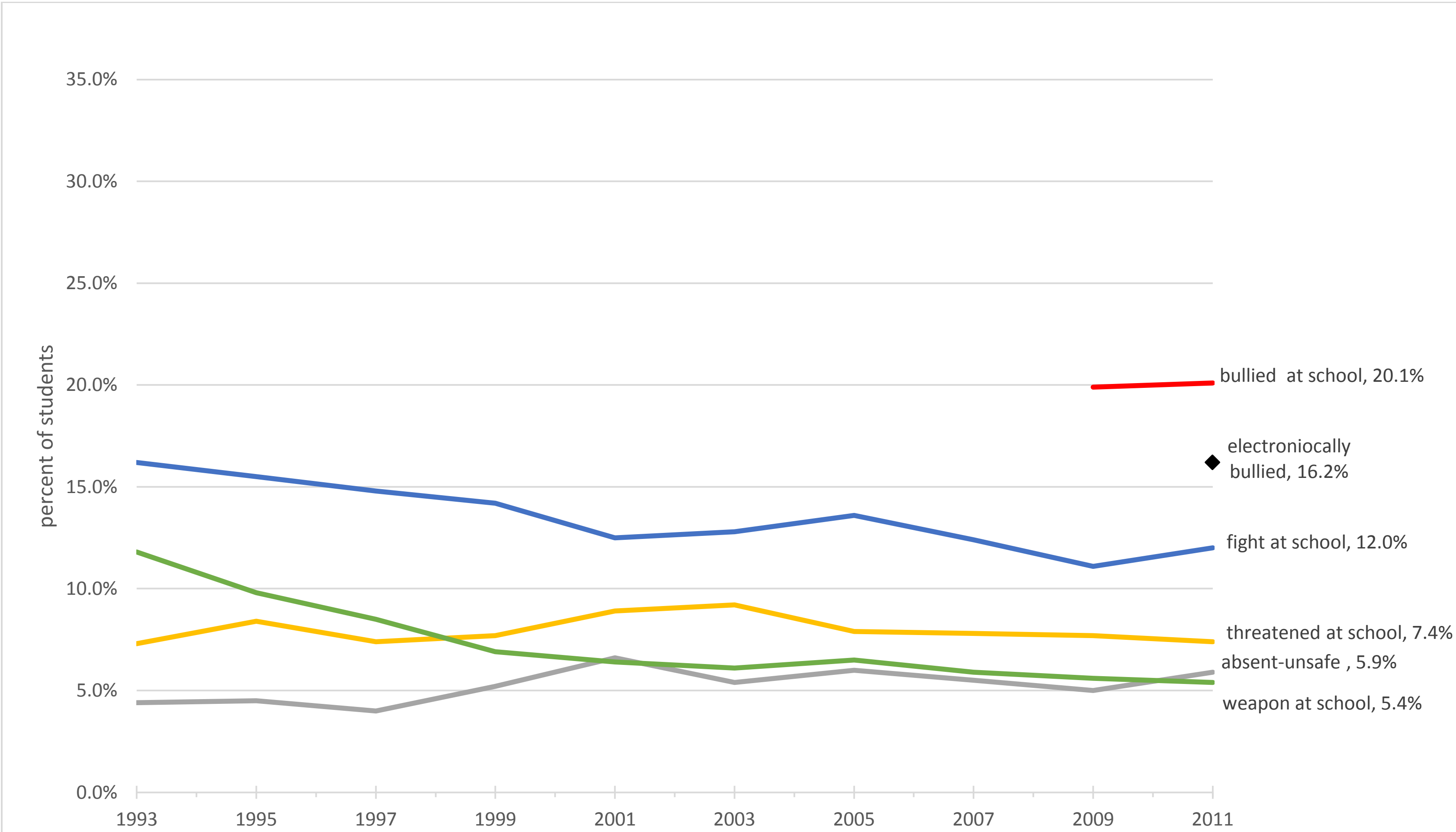
SCS/NCVS sample weights are used. Middle school refers to grades 6-8.

Appendix Table B9. DD Estimates of the Effect of Bullying Laws, YRBSS 1993-2011

Outcomes:	Repeatedly absent-unsafe	Repeatedly fight	Repeatedly threatened	Repeatedly weapon
Policy variables	(1)	(2)	(3)	(4)
ANY Law	-0.0042*	-0.0037	-0.0020	-0.0006
	(0.0023)	(0.0036)	(0.0043)	(0.0029)
Observations	133,552	133,552	133,552	133,552
R-squared	0.0085	0.0241	0.0103	0.0269
MANDATE Law	0.0001	0.0018	-0.0007	0.0059
	(0.0038)	(0.0053)	(0.0068)	(0.0085)
Observations	82,208	82,208	82,208	82,208
R-squared	0.0102	0.0248	0.0110	0.0247
MANDATE-DEADLINE Law	-0.0057	0.0040	-0.0039	0.0039
	(0.0034)	(0.0045)	(0.0039)	(0.0052)
Observations	66,384	66,384	66,384	66,384
R-squared	0.0071	0.0232	0.0104	0.0306

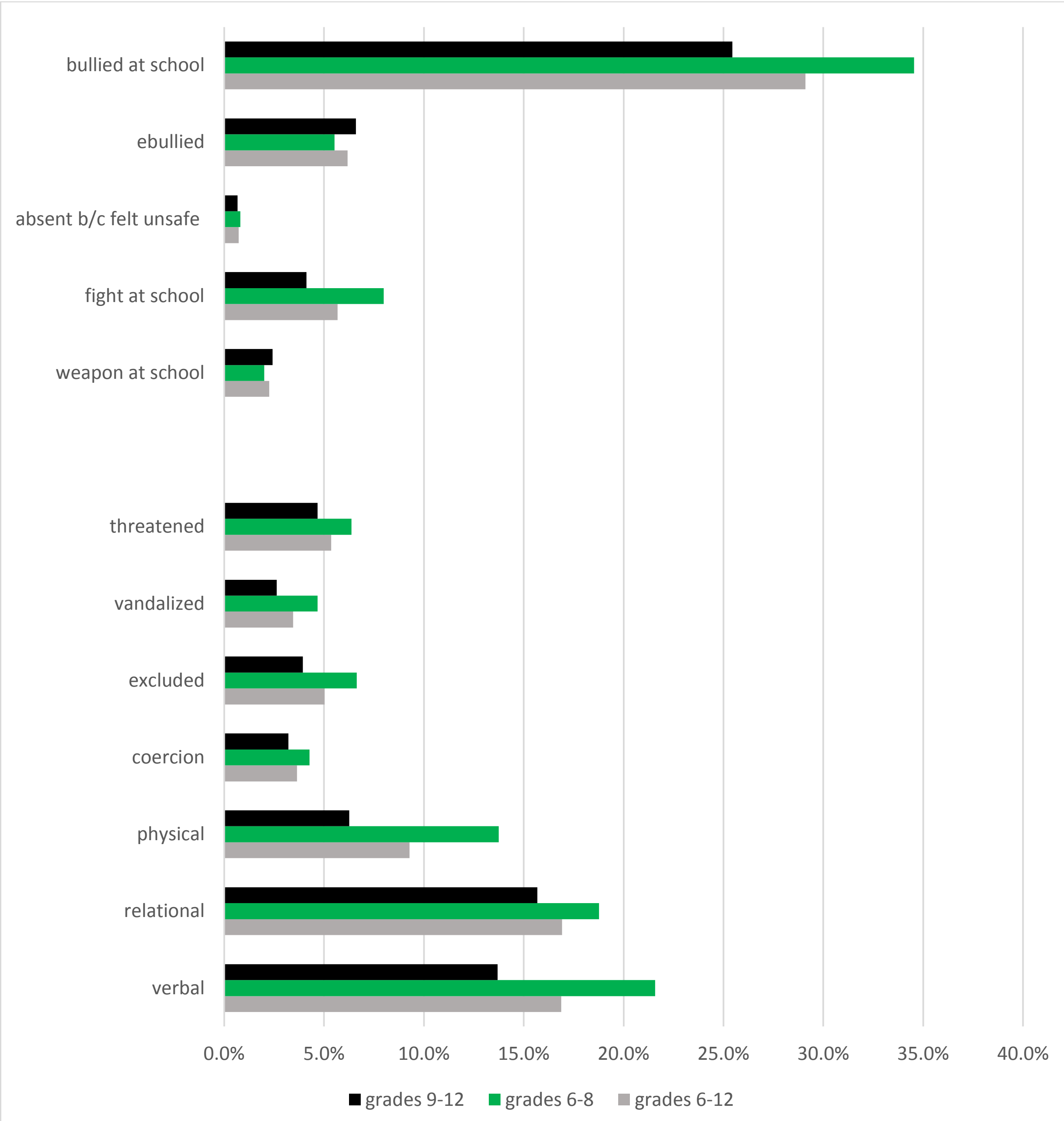
Cluster-robust standard errors in parentheses, clustered by state. *** p<0.01, ** p<0.05, * p<0.10. YRBSS sample weights are used. All regressions include the same covariates as in Table 3..

Appendix Figure B1. National Trends in Bullying-related Experiences among High School Students, YRBSS 1993-2011



Means are calculated using sample weights provided in the YRBSS for years 1993-2011.

Appendix Figure B2. Means of Correlates and Forms of Bullying, by Middle/High school, SCS 2005-2011

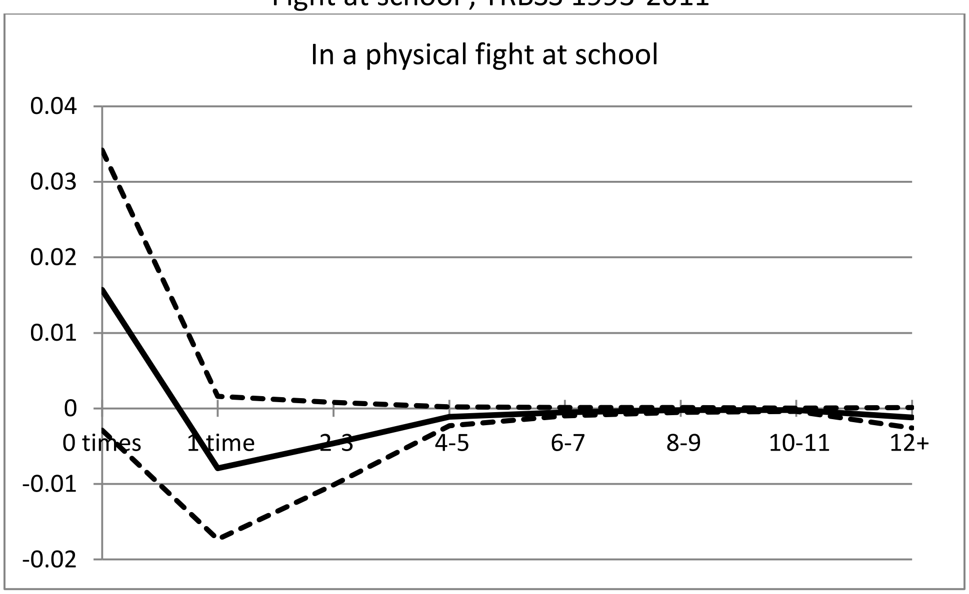


Means are calculated using sample weights provided in the SCS/NCVS for years 2005-2011. Ebullied refers to electronically bullied and the question was included in survey years 2007-2011.

Appendix Figures B3. Marginal Effects of ANY Bullying Law on the Frequency of 'Absent because felt unsafe', YRBSS 1993-2011

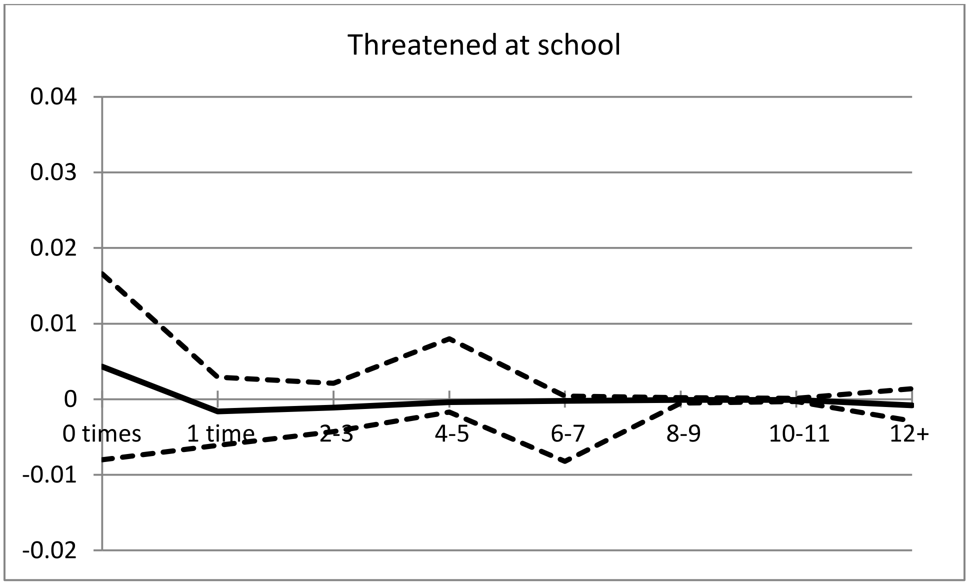
Marginal effects are represented by solid line; 95% confidence interval is represented by the dotted lines. YRBSS sampling weights are used. All marginal effects are statistically significant at 10% level.

Appendix Figures B4. Marginal Effects of ANY Bullying Law on the Frequency of 'Fight at school', YRBSS 1993-2011



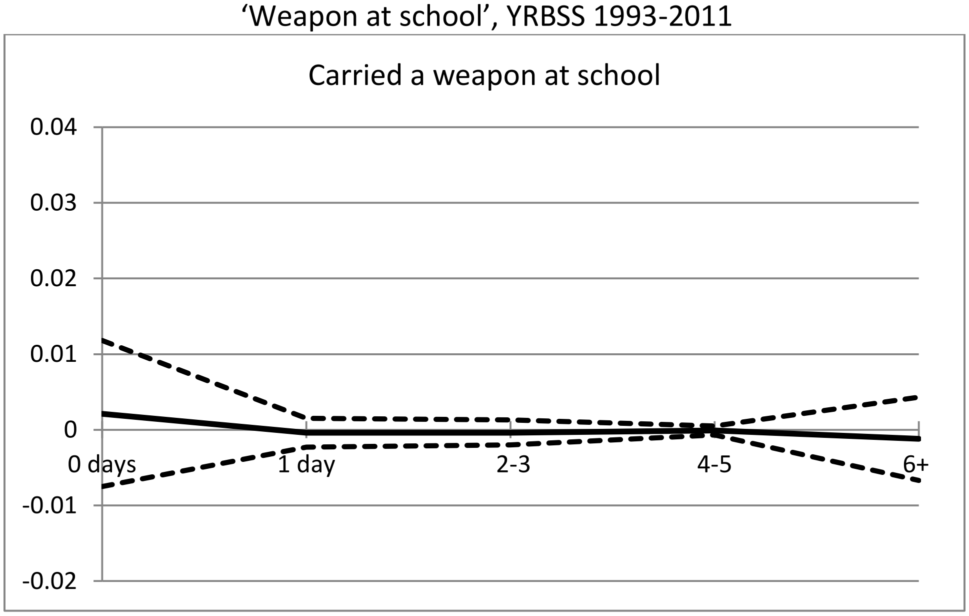
Marginal effects are represented by solid line; 95% confidence interval is represented by the dotted lines. YRBSS sampling weights are used. All marginal effects are statistically insignificant at 10% level except for frequency category 12 or more times, which is statistically significant at 10%.

Appendix Figures B5. Marginal Effects of ANY Bullying Law on the Frequency of 'Threatened at school', YRBSS 1993-2011



Marginal effects are represented by solid line; 95% confidence interval is represented by the dotted lines. YRBSS sampling weights are used. All marginal effects are statistically insignificant at 10% level.

Appendix Figures B6. Marginal Effects of ANY Bullying Law on the Frequency of 'Weapon at school', YRBSS 1993-2011



Marginal effects are represented by solid line; 95% confidence interval is represented by the dotted lines. YRBSS sampling weights are used. All marginal effects are statistically insignificant at 10% level.

C APPENDIX TO CHAPTER III

Appendix Table C1. Examining Key Covariates for SI Participation

Dependent variable:	SI Participation						
Covariates	(1)	(2)	(3)	(4)	(5)	(6)	(7)
=1 if in treatment group	0.3485***			0.2856***		0.2243***	0.2487***
	(0.0954)			(0.0711)		(0.0780)	(0.0827)
SI Participation(Pre-Treat)		2.9497***		2.9481***	2.9362***	2.9346***	2.9499***
		(0.0737)		(0.0737)	(0.0854)	(0.0854)	(0.0942)
High School GPA			0.9103***		0.2379*	0.2407*	0.1872
			(0.1580)		(0.1238)	(0.1236)	(0.1353)
Male							-0.0882
							(0.0890)
Black							0.5415***
							(0.1050)
Asian							0.3871***
							(0.1257)
Other race							0.0334
							(0.1596)
Hispanic							0.4146**
							(0.1684)
Observations	8,879	8,879	6,962	8,879	6,962	6,962	6,171
R-squared	0.0015	0.4412	0.0055	0.4422	0.4482	0.4489	0.4420
Adj. R-squared	0.00140	0.441	0.00534	0.442	0.448	0.449	0.441
AIC	51793.86	46640.35	40199.48	46626.15	36100.28	36093.89	32020.57

Robust standard errors in parentheses, clustered by student id. *** p<0.01, ** p<0.05, * p<0.10.

Appendix Table C2. Examining Key Covariates for Course Grade.

Dependent variable:	Course Grade						
Covariates	(1)	(2)	(3)	(4)	(5)	(6)	(7)
=1 if in treatment group	0.0111						
	(0.0236)						
SI Participation		0.0272***				0.0227***	0.0260***
		(0.0024)				(0.0036)	(0.0037)
SI Participation(Pre-Treat)			0.0983***		0.0523***	-0.0145	-0.0216
			(0.0111)		(0.0119)	(0.0164)	(0.0177)
High School GPA				0.9287***	0.9167***	0.9113***	0.9163***
				(0.0394)	(0.0393)	(0.0393)	(0.0420)
Male							0.1022***
							(0.0302)
Black							-0.3559***
							(0.0360)
Asian							-0.1675***
							(0.0421)
Other race							-0.1639***
							(0.0566)
Hispanic							-0.0504
							(0.0535)
Observations	8,879	8,879	8,879	6,962	6,962	6,962	6,171
Adj. R-squared	-8.79e-05	0.0118	0.00774	0.0948	0.0970	0.102	0.121
AIC	27176.01	27070.13	27106.21	19974.13	19958.13	19923.44	17541.5

Robust standard errors in parentheses, clustered by student id. *** p<0.01, ** p<0.05, * p<0.10.

Appendix Table C3. Summary Statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Grade	8879	2.6914	1.1177	0	4.3
Treatment	8879	0.5039	0.5000	0	1
Number of SI visits	8879	1.9239	4.4741	0	34
Number of SI visits, pre-Treatment	8879	0.3506	1.0075	0	8
HS GPA	6962	3.3640	0.3542	0	4
Male	8879	0.4100	0.4919	0	1
White	8346	0.3091	0.4622	0	1
Black	8346	0.4428	0.4968	0	1
Asian	8346	0.1848	0.3881	0	1
Other	8346	0.0633	0.2435	0	1
Hispanic	8275	0.1132	0.3169	0	1
Age (years)	8879	20.8447	3.8410	14.6	61.7
US citizen	8879	0.9076	0.2895	0	1
Non-citizen	8879	0.0924	0.2895	0	1
Freshman	8879	0.3372	0.4728	0	1
Sophomore	8879	0.3620	0.4806	0	1
Junior	8879	0.1720	0.3774	0	1
Senior	8879	0.1105	0.3135	0	1
Other level	8879	0.0184	0.1342	0	1
Institutional GPA	5692	3.0840	0.6497	0	4.3
Freshman Index	1977	2745.5	229.6	2051	3474
Lives on campus	8879	0.0000	0.0000	0	0
Pell-eligible	8879	0.5986	0.4902	0	1
First-generation college goer	6852	0.2820	0.4500	0	1

Based on full sample before dropping observations due to missing data on key covariates such as HS GPA, sex, and race/ethnicity.

Appendix Table C4. Estimated Per-Session Average Causal Effect of SI Participation on Grade

First-stage					
Dependent variable:	SI Participation				
Covariates	(1)	(2)	(3)	(4)	(5)
=1 if in Treatment group	0.3217***	0.3235***	0.2467***	0.2480***	0.3215***
	(0.1100)	(0.1096)	(0.0827)	(0.0825)	(0.1097)
SI Participation (Pre-Treatment)			2.9630***	2.9543***	
			(0.0940)	(0.0937)	
High School GPA	0.8251***	0.7218***			
	(0.1565)	(0.1570)			
Male		-0.4387***		-0.1134	-0.5383***
		(0.1114)		(0.0834)	(0.1107)
Black		0.6962**		0.5352***	0.6727***
		(0.1294)		(0.1004)	(0.1293)
Asian		0.6337***		0.3893***	0.6437***
		(0.1549)		(0.1206)	(0.1554)
Other race		0.2265		0.0271	0.2032
		(0.2145)		(0.1556)	(0.2149)
Hispanic		0.1485		0.4127**	0.1747
		(0.1903)		(0.1568)	(0.1904)
Observations	6171	6171	6171	6171	6171
Adj. R-squared	0.0056	0.0127	0.4384	0.4412	0.0095
Second-stage					
Dependent variable:	Course Grade				
Covariates	(1)	(2)	(3)	(4)	(5)
SI Participation	0.0366	0.0310	0.0310	0.0222	0.0231
	(0.0798)	(0.0783)	(0.1094)	(0.1074)	(0.0828)
SI Participation (Pre-Treatment)			-0.0190	0.0115	
			(0.3243)	(0.3175)	
High School GPA	0.8853***	0.9088***			
	(0.0768)	(0.0696)			
Male		0.1070**		-0.0215	-0.0227
		(0.0453)		(0.0331)	(0.0536)
Black		-0.3605***		-0.3847***	-0.3848***
		(0.0662)		(0.0697)	(0.0683)
Asian		-0.1725***		-0.1551**	-0.1547**
		(0.0659)		(0.0621)	(0.0706)
Other race		-0.1664***		-0.1946***	-0.1941***
		(0.0588)		(0.0613)	(0.0634)
Hispanic		-0.0492		-0.0137	-0.0148
		(0.0546)		(0.0715)	(0.0575)
Observations	6,171	6,171	6,171	6,171	6,171
Adj. R-squared	0.0940	0.1190	0.0095	0.0316	0.0318

Robust standard errors in parentheses, clustered by student id.*** p<0.01, ** p<0.05, * p<0.10.

Appendix Table C5. Estimated Correlation Between SI Participation and Grade, Control Group

Dependent variable:	Course Grade				
Covariates	(1)	(2)	(3)	(4)	(5)
SI Participation	0.0226***	0.0264***	0.0224***	0.0268***	0.0296***
	(0.0039)	(0.0039)	(0.0058)	(0.0057)	(0.0040)
SI Participation (Pre-Treatment)			0.0264	0.0195	
			(0.0264)	(0.0259)	
High School GPA	0.9131***	0.9169***			
	(0.0555)	(0.0565)			
Male		0.0757*		-0.0440	-0.0455
		(0.0407)		(0.0421)	(0.0420)
Black		-0.4327***		-0.4648***	-0.4658***
		(0.0479)		(0.0499)	(0.0499)
Asian		-0.2055***		-0.2000***	-0.2006***
		(0.0563)		(0.0597)	(0.0597)
Other race		-0.1363*		-0.1408*	-0.1410*
		(0.0729)		(0.0796)	(0.0797)
Hispanic		-0.0461		-0.0088	-0.0099
		(0.0700)		(0.0722)	(0.0722)
Observations	3,093	3,093	3,093	3,093	3,093
R-squared	0.1005	0.1310	0.0105	0.0433	0.0431
Adj. R-squared	0.0999	0.1290	0.0099	0.0411	0.0413

Robust standard errors in parentheses, clustered by student id. *** p<0.01, ** p<0.05, * p<0.10.

Appendix Figure C1. Sample Communications for SI Intervention

Sample Encouragement Message

Subject: TOMORROW @9AM—Go to SI for ECON2106-001

Body: SI review sessions this week for Prof. XYZ's ECON2106-010 meet at:

9:00 – 9:45 AM on Tuesday mm/dd/yy in GCB123
9:00 – 9:45 AM on Wednesday mm/dd/yy in GCB123
9:00 – 9:45 AM on Thursday mm/dd/yy in GCB123

SI—Go, Learn, Get Better Grades!

Why go to SI?
All students—strong, struggling, or in the middle—benefit from going to SI:

- In past semesters, students who attend SI sessions earned, on average, at least ½ a letter grade higher than those students in the course who did not attend SI sessions
- SI helps you use your study time more effectively, so if you use SI properly, you might have more study-free weekends

Please do not reply to this message. If you believe that you received this message in error, please email us at sireviewsession@gsu.edu **with the subject and CRN#** and explain briefly the error.

REFERENCES

Acemoglu, D. and D. Autor. (2015) *Lectures in Labor Economics*. [PDF] Retrieved from http://economics.mit.edu/files/4689.

Allegretto, S., Corcoran, S., & Mishel, L. (2004). How does teacher pay compare? Methodological challenges and answers. Washington, DC: Economic Policy Institute.

Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business & Economic Statistics*, 19(1)

Angrist, J., & Imbens, G. (1995). Two-stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity. Journal of the American statistical Association, 90(430): 431-442.

Angrist, J.D. and J.S. Pischke. (2009) Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press: Princeton, NJ. pp. 11-220, 293-325.

Ammermueller, A. (2012). Violence in European schools: A Widespread Phenomenon that Matters for Educational Production. *Labour Economics.* 19(6): 908-922.

Autor, D., Levy, F., & Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. Quarterly Journal of Economics, 11(4), 1279-1333.

Barron, D., & West, E. (2013). The financial costs of caring in the British labour market: Is there a wage penalty for workers in caring occupations? British Journal of Industrial Relations, 51(1), 104-123.

Bayard, K., Hellerstein, J., Neumark, D., & Troske, K. (2003). New evidence on sex segregation and sex differences in wages from matched employee-employer data. Journal of Labor Economics, 21(4), 887-922.

Becker, G. (1975). *Human Capital* (2nd ed.) Chicago, IL: The University of Chicago Press.

Bender, K., & Heywood, J. (2012). Trends in the relative compensation of state and local employees. In D. Mitchell (Ed.), Public jobs and political agendas: The public sector in an era of economic stress (pp. 133-166). Champaign, IL: Labor and Employment Relations Association.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004) How Much Should We Trusted Difference-in-Difference Estimates? *Quarterly Journal of Economics.* 119(1): 249-275.

Bettinger, E., & Slonim, R. (2007). Patience among children. *Journal of Public Economics*, *91*(1), 343-363.

Blackburn, M. (2007). Estimating wage differentials without logarithms. Labour Economics, 14(1), 73-98.

Blanc, R. A., DeBuhr, L. E., & Martin, D. C. (1983). Breaking the attrition cycle: The effects of supplemental instruction on undergraduate performance and attrition. *The Journal of Higher Education*, 80-90.

Blanc, R., & Martin, D. C. (1994). Supplemental instruction: increasing student performance and persistence in difficult academic courses. *Academic Medicine*, 69(6), 452-4.

Bollinger, C.,& Hirsch, B. (2006). Match bias from earnings imputation in the current population survey: The case of imperfect matching. Journal of Labor Economics, 24(3), 483-519.

Bollinger, C., & Hirsch, B. (2013). Is earnings nonresponse ignorable? Review of Economics and Statistics, 95(2), 407-416.

Borghans, L., ter Weel, B., & Weinberg, B. (2008). Interpersonal styles and labor market out-comes. Journal of Human Resources, 43(4), 815-858.

Borghans, L., ter Weel, B., & Weinberg, B. (2014). People skills and the labor market outcomes of underrepresented groups. Industrial and Labor Relations Review, 67(2), 287-334.

Borjas, G. (2013). Labor economics (6th ed.). New York, NY: McGraw-Hill Irwin.

Brown, S. and Taylor, K. (2008) Bullying, Education, and Earnings: Evidence from the National Child Development Study. *Economics of Education Review.* 27: 387-401.

Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. Journal of Human Resources, 33(1), 88-126.

Cameron, S. V., & Heckman, J. J. (2001). The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy*, 109(3), 455-499.

Cameron, C. and D. Miller. (2015) A Practintioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*. 50(2): 317-372.

Card, D. (1996). The effect of unions on the structure of wages: A longitudinal analysis. Econometrica, 64(4), 957-979.

Card, D., & Rothstein, J. (2007). Racial segregation and the black–white test score gap. *Journal of Public Economics*, 91(11), 2158-2184.

Castillo, M., Ferraro, P. J., Jordan, J. L., & Petrie, R. (2011). The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics*, 95(11), 1377-1385.

Center for Disease Control (CDC) website for YRBSS. Accessed 8 June 2013. http://www.cdc.gov/healthyyouth/yrbs/index.htm.

Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association, 74(368), 829-836.

Craig, W., Yossi Harel-Fisch, Haya Fogel-Grinvald, Suzanne Dostaler, Jorn Hetland, Bruce Simons-Morton, Michal Molcho, Margarida Gaspar de Mato, Mary Overpeck, Pernille Due, William Pickett, the HBSC Violence & Injuries Prevention Focus Group, the HBSC Bullying Writing Group (2009). A Cross-national Profile of Bullying and Victimization Among Adolescents in 40 Countries. *International Journal of Public Health*, 54(2), 216-224.

Dake, J. A., Price, J. H., & Telljohann, S. K. (2003). The Nature and Extent of Bullying at School. *Journal of School Health*, 73(5), 173-180.

DeVoe, J.F., and Bauer, L. (2010). Student Victimization in U.S. Schools: Results From the 2007 School Crime Supplement to the National Crime Victimization Survey (NCES 2010-319). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office, accessed 4-21-2015 http://nces.ed.gov/pubs2010/2010319.pdf.

Diagne, D. (2009) School Violence: Evidence from the Economics Literature and Related Disciplines. *Revue Suisse des Sciences de L'education* 31(1): 135-150.

Dobbie, W., & Fryer Jr, R. G. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, 158-187.

Drydakis, N. (2013) Bullying at School and Labor Market Outcomes. *IZA Discussion Paper No.7432.*

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics*, *4*: 3895-3962.

England, P. (2005). Emerging theories of care work. Annual Review of Sociology, 31, 381-399.

England, P., Budig, M., & Folbre, N. (2002). Wages of virtue: The relative pay of care work. Social Problems, 49(4), 455-473.

England, P., & Folbre, N. (1999). The cost of caring. ANNALS of the American Academy of Political and Social Science, 561,39-51.

Eriksen, T. L. M., Nielsen, H. S., & Simonsen, M. (2013). Bullying in Elementary School. *Journal of Human Resources*. 49(4): 839-871.

Fehr, E., & Ga chter, S. (2000). Fairness and retaliation: The economics of reciprocity. Journal of Economic Perspectives, 14(3), 159-181.

Frisén, A., Hasselblad, T., & K. Holmqvist. (2012). What actually makes bullying stop? Reports from former victims. *Journal of Adolescence*, 35(4): 981-990.

Folbre, N. (2006). Demanding quality: Worker/consumer coalitions and “high road” strategies in the care sector. Politics and Society, 34(1), 11-31.

Folbre, N. (2008). When a commodity is not exactly a commodity. Science, 319, 1769-1770.

Folbre, N. (2012). Should women care less? Intrinsic motivation and gender inequality. British Journal of Industrial Relations, 50(4), 597-619.

Folbre, N., & Nelson, J. (2000). For love or money –Or both? Journal of Economic Perspectives, 14(4), 123-140.

Fryer Jr, R. G., & Levitt, S. D. (2013). Testing for racial differences in the mental ability of young children. *American Economic Review*, 103(2), 981-1005.

Georgia State University website for Supplemental Instruction. Accessed 1 September 2012. http://success.students.gsu.edu/success-programs/supplemental-information/.

Gerber,A. and D. Green (2012) *Field Experiments: Design, Analysis, and Interpretation.* W.W. Norton& Company: New York, NY.

Glew, G. M., Fan, M. Y., Katon, W., Rivara, F. P., & Kernic, M. A. (2005). Bullying, Psychosocial Adjustment, and Academic Performance in Elementary school. *Archives of Pediatrics & Adolescent Medicine*, 159(11), 1026-1031.

Gittleman, M., & Pierce, B. (2011). Compensation for state and local government workers. Journal of Economic Perspectives, 26(1), 217-242.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, 191-209.

Greene, M. B. (2003) High School Students Are Also Adversely Affected by Bullying. *Archives of Pediatrics & Adolescent Medicine.* 157(11): 1134.

Grogger, J. (1997). Local Violence and Educational Attainment. *Journal of Human Resources*. 32(4): 659-682.

Harrison, G. & List. J. (2004) Field Experiments. *Journal of Economic Literature.* 42(4):1009-1055.

Health Behaviour in School-aged Children (HBSC) Study: International Report From the 2001/2002 survey. http://www.hbsc.org, accessed 1 July 2013.

Health Behaviour in School-aged Children (HBSC) Study: International Report From the 2009/2010 survey. http://www.hbsc.org, accessed 1 July 2013

Heckman, J., & Kautz, T. (2012). Hard evidence on soft skills. Labour Economics, 19(4), 451-464.

Heckman, J. (1997) Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *Journal of Human Resources.* 32(3): 441-460.

Hersch, J. (1998). Compensating differentials for gender-specific job injury risks. American Economic Review, 88(3), 598-607.

Hirsch, B., & Macpherson, D. (2014). Union membership and earnings data book: Compilations from the current population survey (2014 ed.). Arlington, VA: The Bureau of National Affairs.

Hirsch, B., & Schumacher, E. (2012). Underpaid or overpaid? Wage analysis for nurses using job and worker attributes. Southern Economic Journal, 78(4), 1096-1119.

Hoxby, C., & S. Turner. (2015) What High-Achieving Low-Income Students Know about College. *American Economic Review*, 105(5): 514-17.

Huynh, K. P., Jacho-Chávez, D. T., & Self, J. K. (2010). The efficacy of collaborative learning recitation sessions on student outcomes. *The American Economic Review*, 100(2): 287-291.

Hwang, H., Reed, W. R., & Hubbard, C. (1992). Compensating wage differentials and unobserved productivity. Journal of Political Economy, 100(4), 835-858.

Imai, K. & M. Ratkovic (2013). Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation. *Annals of Applied Statistics*, 7(1): 443-470.

Joinson, A. N., & Reips, U. D. (2007). Personalized salutation, power of sender and response rates to Web-based surveys. *Computers in Human Behavior*, 23(3): 1372-1383.

Kochenderfer, B. J., & Ladd, G. W. (1996). Peer Victimization: Cause or Consequence of School Maladjustment? *Child Development*. 67(4): 1305-1317.

Koenker, R. (2005). Quantile regression. Cambridge: Cambridge University Press.

Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2015). Behavioral economics of education: Progress and possibilities (IZA DP No. 8853), Institute for the Study of Labor.

Le, A., P. Miller, A. Heath, and N. Martin. (2005). Early Childhood Behaviours, Schooling and Labour Market Outcomes: Estimates From a Sample of Twins. *Economics of Education Review*. 24(1): 1-17.

Leonhardt, D. (2015) College for the Masses. *The New York Times*, 24 April 2015. Accessed on 16 June 2015 at http://www.nytimes.com/2015/04/26/upshot/college-for-the-masses.html?_r=0&abt=0002&abg=0

Lewin, D., Keefe, J., & Kochan, T. (2012). The new great debate about unionism and collec¬tive bargaining in U.S. state and local governments. Industrial and Labor Relations Review, 65(4), 749-778.

Lewis, D., M. O'Brien, S. Rogan, and B. Shorten. (2005) "Do Students Benefit From Supplemental Instruction? Evidence from a First Year Statistics Subject in Economics and Business. Working Paper? University of Wollongong, Austrailia.

Linneman, P., & Wachter, M. (1990). The economics of federal compensation. Industrial Relations, 29(1), 58-76.

Lochner, L., & Moretti, E. (2015). Estimating and Testing Models With Many Treatment Levels and Limited Instruments. Review of Economics and Statistics. 97(2): 387-397.

Loviscek, A., and Cloutier, N. (1997) Supplemental Instruction and the Enhancement of Student Performance in Economics Principles. *The American Economist*. 41(2): 70-76.

Ludsteck, J. (2014). The impact of segregation and sorting on the gender wage gap: Evidence from German linked longitudinal employer-employee data. Industrial and Labor Relations Review, 67(2), 362-393.

Macpherson, D., & Hirsch, B. (1995). Wages and gender composition: Why do women's jobs pay less? Journal of Labor Economics, 13(3), 426-471.

Madrian, B., & Lefgren, L. (2000). An approach to longitudinally matching Current Population Survey (CPS) respondents. Journal of Economic and Social Measurement, 26(1), 31-62.

Malm, J., Bryngfors, L., & Morner, L. L. (2011). Supplemental Instruction: Whom Does It Serve?. *International Journal of Teaching and Learning in Higher Education*, 23(3), 282-291.

Manning, A. (2003). Monopsony in motion: Imperfect competition in labor markets. Princeton, NJ: Princeton University Press.

Morgan, S.L. and Winship, C. (2007). Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge University Press: New York, NY.

Nansel, T., Overpeck, M., Hyanie, D., Ruan, J., and P. Scheidt. (2003) Relationships Between Bullying and Violence Among US Youth. *Archives of Pediatrics & Adolescent Medicine.* 157(4): 348-353.

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying Behaviors Among US Youth. *JAMA: the Journal of the American Medical Association 285*(16): 2094-2100.

Ogden, P., Thompson, D., Russell, A., & Simons, C. (2003). Supplemental Instruction: Short-and Long-Term Impact. *Journal of Developmental Education*, 26(3): 2-8.

Olweus, D. (1997) Bully/Victim Problems in School: Facts and Intervention. *European Journal of Psychology of Education.* 12(4): 495-510.

Olweus, D. (2003) A Profile of Bullying at School. *Educational Leadership*. 3: 12-17.

O*NET OnLine Help. Scales, ratings, and standardized scores. Retrieved from http://www. Onetonline.org/help/online/scales. Accessed on October 10, 2013.

Podgursky, M., Monroe, R., & Watson, D. (2004). The academic quality of public school teachers: An analysis of entry and exit behavior. Economics of Education Review, 23(5), 507-518.

Powdthavee, N. (2012) Resilience to Economic Shocks and the Long Reach of Childhood Bullying. *IZA Discussion Paper No. 6645.*

Ramirez, Gen M. (1997) Supplemental Instruction: The Long-Term Impact. *Journal of Developmental Education,* 21(1): 2-10.

Raynor, S. & A. Wylie. (2012). Presentation and management of school bullying and the impact of anti-bullying strategies for pupils: A self-report survey in London schools. *Public Health*, 126(9): 782-789.

Scafidi, B., Sjoquist, D., & Stinebrickner, T. (2006). Do teachers really leave for higher paying jobs in alternative occupations? Advances in Economic Analysis and Policy, 6(1),1-42.

Schumacher, E., & Hirsch, B. (1997). Compensating differentials and unmeasured ability in the labor market for nurses: Why do hospitals pay more? Industrial and Labor Relations Review, 50(4), 557-579.

Spewak, D. (2015) "109 Schools in WNY Reported Zero Bullying Incidents" Accessed 16 February 2015 http://www.wgrz.com/story/news/2015/02/01/identifying-the-bullies/22692143/

Srabstein, J., Berkman, B., and Pyntikova, E. (2008) Antibullying Legislation: A Public Health Perspective. *Journal of Adolescent Health*. 42: 11-20.

Stinebrickner, R. and T. Stinebrickner. (2008) The Effect of Credit Constraints on the College Drop-Out Decision: A Direct Approach Using a New Panel Study. *American Economic Review*, 98(5): 2163-84.

Stock, W. A., Ward, K., Folsom, J., Borrenpohl, T., Mumford, S., Pershin, Z., Carriere, D. & Smart, H. (2013). Cheap and Effective: The Impact of Student-Led Recitation Classes on Learning Outcomes in Introductory Economics. *The Journal of Economic Education*, 44(1): 1-16.

Stuart-Cassell, V., A. Bell, and J.F. Springer (2011) Analysis of State Bullying Laws and Policies. Report submitted to US Department of Education, Office of Planning, Evaluation, and Policy Development.

Thaler, R., & Rosen, S. (1976). The value of saving a life: Evidence from the labor market. In N. Terleckj (Ed.), Household production and consumption (pp. 265-298). New York, NY: Columbia University Press.

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*, 1(1), 91-136.

Ttofi, M. & D. Farrington. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, 7(1): 27-56.

United States Department of Justice. Bureau of Justice Statistics. National Crime Surveys: Crime School Supplement, 1989 [Computer file]. ICPSR09394-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1994. http://doi.org/10.3886/ICPSR09394.v2

U.S. Dept. of Justice, Bureau of Justice Statistics. National Crime Surveys: Crime School Supplement, 1995 [Computer file]. Conducted by U.S. Dept. of Commerce, Bureau of the Census. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1998. doi:10.3886/ICPSR06739.v1

United States Department of Justice. Bureau of Justice Statistics. National Crime Victimization Survey: School Crime Supplement, 1999 [Computer file]. ICPSR03137-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2001. doi:10.3886/ICPSR03137.v1

United States Department of Justice. Bureau of Justice Statistics. National Crime Victimization Survey: School Crime Supplement, 2001 [Computer file]. ICPSR03477-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2002. doi:10.3886/ICPSR03477.v1

U.S. Dept. of Justice, Bureau of Justice Statistics. National Crime Surveys: Crime School Supplement, 2003 [Computer file]. Conducted by U.S. Dept. of Commerce, Bureau of the Census. ICPSR04182.v1. Ann Arbor, MI: Inter-university Consortium of Political and Social Research [producer and distributor], 2005-07-29. doi:10.3886/ICPSR04182.v1

U.S. Dept. of Justice, Bureau of Justice Statistics. National Crime Surveys: Crime School Supplement, 2005 [Computer file]. Conducted by U.S. Dept. of Commerce, Bureau of the Census. ICPSR04429-v2. Ann Arbor, MI: Inter-university Consortium of Political and Social Research [producer and distributor], 2008-04-30. doi:10.3886/ICPSR04429.v2

United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. National Crime Victimization Survey: School Crime Supplement, 2007 [Computer file]. ICPSR23041-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-03-26. doi:10.3886/ICPSR23041.v1

United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. National Crime Victimization Survey: School Crime Supplement, 2009 [Computer file]. ICPSR28201-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-01-21. doi:10.3886/ICPSR28201.v1

United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. National Crime Victimization Survey: School Crime Supplement, 2011 [Computer file]. ICPSR33081-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-03-26. doi:10.3886/ICPSR33081.v1

United States Government Accountability Office (2012) School Bullying: Extent of Legal Protections for Vulnerable Groups Needs to Be More Fully Assessed. *Report to Congressional Requestors*. GAO-12-349, May 2012.

Van der Klaauw, W. (2002) Estimating The Effect of Financial Aid Offers On College Enrollment: A Regression–Discontinuity Approach. *International Economic Review*, 43(4): 1249-1287.

Van Houtven, C., Coe, N., & Skira, M. (2013). The Effect of Informal Care on Work and Wages. Journal of Health Economics, 32(1), 240-252.

Vara, V. (2015) Is College the New High School? *The New Yorker,* 13 January 2015. Accessed 16 June 2015 http://www.newyorker.com/business/currency/college-new-high-school

Varhama L.M., and Bjorkqvist K. (2005) Relation Between School Bullying During Adolscence and Subsequent Long-term Unemployment in Adulthood in a Finnish Sample. *Psychological Reports.* 96(2): 269-272.

Waddell, G.R. (2006) Labor-Market Consequences of Poor Attitude and Low Self-Esteem in Youth. *Economic Inquiry*. 44(1): 69-97.

Wang, J., Iannotti, R. J., & Nansel, T. R. (2009). School Bullying Among Adolescents in the United

States: Physical, Verbal, Relational, and Cyber. *Journal of Adolescent Health.* 45(4): 368-375.

Webber, D. (2013). Firm-level monopsony and the gender pay gap. IZA Discussion Paper No. 7343, April.

Wilson, B. and S. Rossig. (2013) "Does Supplemental Instruction for Principles of Economics Help Close the Gap for Traditionally Underrepresented Minorites?" Unpublished manuscript. Humboldt State University.

VITA

Julia Manzella was born and raised in Western New York State, USA. She holds a Bachelor of Arts degree in Mathematics from the State University of New York College at Geneseo and Master's of Arts degree in Economics from the University of Arizona.

Julia began the Doctoral program in Economics at Georgia State University in 2010 to study Labor Economics and Applied Econometrics. She was actively engaged in the academic community. Julia was selected by administration in the Andrew Young School of Policy Studies from a group of volunteers to be a representative in external and internal organizations. From 2012 to 2013, she served as a member of the Policy Review Task Force for College Completion, a committee comprised of faculty, staff, and students from institutions across the state put together by the Board of Regents of the University System of Georgia. From 2014 to 2015, she was a member of the Graduate Student Alliance, a campus-wide student organization aiming to facilitate communications between university administrators and graduate students.

Julia has been the recipient of several internal awards. In 2011 she received the Carole Keels Scholarship in Economics in recognition of her significant career experience. In 2014 she received the Dan E. Sweat Dissertation Fellowship for outstanding research addressing urban, community, or education policy issues. In 2015 she received the George Malanos Economics Scholarship in recognition of her commitment to the exchange of ideas and the creation of a community of scholars.

Prior to studying economics, Julia had a successful career teaching a wide range of mathematics courses at the high school and community college levels. As a graduate student, she was invited to give guest lectures in undergraduate (Principles of Microeconomics) and graduate (Econometrics—PhD level) courses.

Inspired by disparities and inefficiencies she observes in her professional and personal experiences, she employs a variety of economic tools in her empirical research in order to better understand why some individuals flourish (and others do not) in the labor market and how policy might be used to increase the likelihood of success. Julia's research has been published in *Research in Labor Economics*. She has presented her research in many different settings: academic seminars, economics conferences, government agencies, and internal briefings. Notably, her sole authored paper on the impact of state bullying laws has been accepted to the 2015 Society of Labor Economists/European Association of Labour Economists World Conference. She has also served as a referee for the journal *Contemporary Economic Policy*.

Julia received her Doctor of Philosophy degree in Economics from Georgia State University in August, 2015. She looks forward to embarking on a new career as a Labor Economist.