

2021

Causal Determinants of College Success

Benjamin Blemings

West Virginia University, bbleming@mix.wvu.edu

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>



Part of the [Economics Commons](#)

Recommended Citation

Blemings, Benjamin, "Causal Determinants of College Success" (2021). *Graduate Theses, Dissertations, and Problem Reports*. 8079.

<https://researchrepository.wvu.edu/etd/8079>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

2021

Causal Determinants of College Success

Benjamin Blemings

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

 Part of the Economics Commons

Causal Determinants of College Success

Benjamin Tyler Blemings

**Dissertation submitted to the
John Chambers College of Business and Economics
at West Virginia University**

**in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Economics**

**Adam Nowak, Ph.D., Committee Chairperson
Daniel Grossman, Ph.D.
Bryan McCannon, Ph.D.
Scott Cunningham, Ph.D.
Brad Humphreys, Ph.D.
Department of Economics**

**Morgantown, West Virginia
2021**

**Keywords: College Success, Causal Inference, Human Capital, Merit Aid, Peer Effects,
Alcohol Access, Crime, College Football, Promise Scholarship, Non-cognitive Ability**

Copyright © 2021 Benjamin Tyler Blemings

Abstract

Causal Determinants of College Success

Benjamin Tyler Blemings

The first essay focuses on alcohol access in college, which is an impediment to many students' success. Alcohol sales to general seating at college football games has been broadly adopted despite results suggesting greater alcohol availability increases crime and college alcohol abuse. Using differences-in-differences, this paper estimates how allowing alcohol sales in college football general seating sections affects crime. Reductions in crime are found after sales are allowed. After sales begin, those under the age of 21 are still not legally permitted to purchase in-stadium alcohol, but are still affected by changes in other stadium policies or law enforcement strategies. Using this group to address unobserved confounders suggests the causal effect of increased legal alcohol access at games is reduced crime on home game days. Results suggest substitution effects matter for the relationship between alcohol access and criminal behavior, identifying a new empirical nuance to be heeded when making alcohol policies.

The second essay focuses on financial aid for college. Merit-based financial aid programs are amongst the largest sources of financial aid for college, yet estimated effects may be biased by students sorting into treatment due to known academic cutoffs. This paper estimates how likely students are to meet minimum academic renewal requirements, in a state where sorting into treatment has been shown, by using an unmanipulated birthdate cutoff created by compulsory kindergarten entry laws. Fuzzy regression discontinuity shows being born slightly after the cutoff, 14 years before the college program begins, leads to a 49.6 percentage point (pp) increase in scholarship receipt. Difference in discontinuity estimates, accounting for potential relative age and season of birth confounding, find students are 24 pp more likely to exceed GPA and completed credits renewal requirements. The results suggest GPA or credits thresholds are equally effective, advancing the literature on behavioral aspects of large merit-aid programs and informing their design.

The third essay focuses on peer groups and non-cognitive abilities. Employers value non-cognitive abilities as much as cognitive ones and non-cognitive abilities have significant bearing on education, labor market, health, and macroeconomic outcomes. This paper estimates how peers affect non-cognitive skill formation, including communication, leadership, and attitude. Students at the National Outdoor Leadership School are conditionally randomly assigned to groups, meaning they are randomly exposed to different levels of peer ability. Being assigned to peers with higher self-rated non-cognitive ability causes a reduction in self-rated communication and leadership by 0.1 standard deviations, but not attitude. The largest results are from females rating themselves lower in communication and leadership. This work suggests strong peers are not an effective way to bolster non-cognitive ability for college-aged individuals and that peers may contribute to less women in leadership positions.

Acknowledgments

I thank my family. My dad for consistently encouraging me to challenge myself academically and intellectually. My mom and sister for always being supportive, even while rolling their eyes at my economics-based comments. I thank my dog, Sasha, who was the go to family member for my mental health. My extended family on both my mom and dad's sides are also a network of top-tier quality.

I thank my academic mentors, who are too numerous to count. My first exposure to economics was Dr. Steve Elliot who picqued my interest in economics. My undergraduate thesis committee, Dr. Charles Moul, Dr. Deborah Fletcher, and Dr. William Even for believing in me. I thank my many mentors at West Virginia University including Dr. Daniel Grossman, Dr. Adam Nowak, Dr. Bryan McCannon, Dr. Brad Humphreys, and Dr. Greg DeAngelo for pushing me and always being willing to discuss research. I thank Dr. Josh Hall for financial and teaching support, reading several drafts, and forcing us to learn LaTeX. Throughout the program, I met other prestigious economists, but the two who I learned the most from are Dr. Scott Cunningham and Dr. Mark Hoekstra.

I thank my classmates and friends at the undergraduate and graduate level. Without Jessica Gunnerson, intermediate macroeconomic theory and introductory econometrics would have been more impossible and lonely. My first year crew of Eduardo Minuci, Alexandre Scarciofollo, and Daniel Bonneau were a terrific support group. Without Margaret Bock, comprehensive exams could have ended my Ph.D. journey and GIS would feel more like Pig Latin than Spanish.

Many other university affiliates were invaluable. I immensely enjoyed spending time with our extremely helpful administrative assistant, Karen Smith. She enjoyed hanging out with me only because I was lucky enough to be affiliated with the Hearts of Gold service dog training program at WVU. I also owe a thank you to Elizabeth Reynolds who provided administrative data that I would not have been able to access otherwise.

Contents

Abstract	ii
Acknowledgments	iii
List of Figures	vii
List of Tables	x
1 Not So Fast: Alcohol Sales to College Football General Seating Sections Reduces Crime	1
1.1 Data	5
1.1.1 Alcohol Sales	6
1.1.2 Descriptive Statistics	6
1.2 Method	7
1.2.1 Selecting a Control Group	9
1.2.2 Identification	10
1.3 Results	11
1.3.1 Home and Away Game Premiums in Pre and Post Periods	11
1.3.2 Difference in Differences Results	12
1.3.3 Inference Robust to Few (Treated) Clusters	14
1.3.4 Event-Study	15
1.3.5 Investigating Omitted Variable Bias	16

1.4	Using Age-Based Variation to Address Omitted Variables	18
1.4.1	Clearance Rate at Legal Drinking Age Cutoff	18
1.4.2	Effect of Alcohol Access on Crime, Using Variation From Age	19
1.4.3	Effect of Alcohol Access on Victimization, Using Variation From Age	21
1.5	Conclusion	21
2	College Merit-Aid and Academic Renewal Thresholds: Evidence from Interactions of Discontinuous Rules	33
2.1	Related Literature	39
2.1.1	Behavior Responses to Academic Renewal Thresholds	39
2.1.2	Research Designs	40
2.1.3	Previous Strategies for Addressing Academic Score Manipulation	41
2.2	Data	42
2.2.1	Outcomes: Academic Performance Based Renewal	43
2.3	Kindergarten Cohort, College Cohort, and the Promise Scholarship	44
2.3.1	Scholastic and Birthdate Cutoff Comparisons	48
2.4	Internal Validity of Research Design	49
2.4.1	First Stage	49
2.4.2	Density Around ACT Composite Score and Birthdate Cutoffs	52
2.4.3	Pre-Treatment Covariate Balance	55
2.5	Methods	58
2.5.1	Identification	59
2.6	Main Results	60
2.6.1	Reduced Form Effects	61
2.6.2	Local Average Treatment Effects from Fuzzy Regression Discontinuity	62
2.6.3	Difference in Discontinuity Estimates	64
2.6.4	Heterogeneous Results	66
2.7	Conclusion	67

3	The Effects of Random Peers on Noncognitive Abilities: Evidence from the National Outdoor Leadership School	80
3.1	The National Outdoor Leadership School	84
3.2	Data	85
3.2.1	Non-Cognitive Skills	86
3.2.2	Descriptive Statistics	87
3.3	Method	89
3.3.1	Peer Effect Specification	89
3.3.2	Identification	90
3.4	Results	92
3.4.1	Binned Scatter Plots of Peer Ability on Individual Ability	92
3.4.2	OLS Regressions of Peer Effects on Noncognitive Abilities	93
3.4.3	Non-Linear Peer Effects	95
3.4.4	Marginal Effects	96
3.4.5	Heterogeneous Effects	96
3.5	Conclusion	99
	Appendix	123
A	Essay 1	123
A.1	Additional Tables and Figures	123
B	Essay 2	135
B.1	Additional Tables and Figures	135
B.2	Compulsory Kindergarten Entry Laws	150
B.3	Threats to Identification	150
C	Essay 3	154
C.1	Additional Tables and Figures	154
C.2	Principal Components Analysis	170

List of Figures

1.1	Years Teams in Sample Changed to Sales	23
1.2	Gameday Crime Premiums for Home Game Versus Away Game Weeks	25
1.3	Randomization Inference, Distribution of Z-Statistics From Home X Sales From Equation 1.1 Using Randomized Treatment Across Seasons and All Teams in Sample	27
1.4	Time-Varying Treatment Effects of Sales on Home Game Crime Compared to Away Game Crime	28
1.5	How Enforcement Changes After Sales at the Legal Alcohol Access Age Threshold	30
2.1	Illustration of Scholarship Differentially Assigned by Small Changes in Birth Date, Compulsory Kindergarten Entry Date Cutoff Laws, and Grade Progression	69
2.2	Eligibility Measures and Scholarship Receipt	70
2.3	Kindergarten Entry Date Cutoffs on Cohort Entry Year and Scholarship Receipt	71
2.4	Comparing Academic and Birthdate Running Variables' Continuity Around Cutoffs	72
2.5	Effect of Being Born After Birthdate Cutoffs on Likelihood of Exceeding Renewal Thresholds in Freshman Year, Academically Eligible Students Only	74
2.6	Coefficient Estimates from Difference in Discontinuity on Freshman Outcomes	77
2.7	Heterogeneity by Academic Credentials	78

2.8	Coefficient Estimates from Difference in Discontinuity on 3 Year Outcomes, by Gender	79
3.1	The Effect of Pre-Course Peer Average Composite Ability on Post-Course Own Composite Ability	104
3.2	Marginal Effects From Non-Parametric Kernel Regressions	106
A.1	Time-Varying Difference in In-Stadium Legal Alcohol Availability by Age Group	124
A.2	Trends in Crime, Home-Away	127
A.3	Examining If Age-Based Variation Feasibly Controls for Enforcement Bias; Density of Having a Report	134
B.1	Merit-Aid Adoption by State and Time	136
B.2	Trends in Cost of College and West Virginia Promise Scholarship Program	137
B.3	1st Year Renewal Likelihood by Cohort	139
B.4	Effect of Birth Month on Years of Scholarship	140
B.5	Relationship between Birth Date and Years of Scholarship Received, Academically Eligible	141
B.6	Pre-Treatment Covariate Continuity at Date Cutoff	142
B.7	Tests of Gender Continuity at Date Cutoff	143
B.8	Regression Discontinuity Identification Assumption: Local Smoothness of Potential Outcomes	144
B.9	Reduced Form Effect of Birthdate on Likelihood of Exceeding Renewal Thresholds in Freshman Year, Academically Ineligible Students	145
B.10	GPA Estimates by Year by Gender	148
B.11	Credits Estimates by Year by Gender	149
C.1	Survey Taken by NOLS Students	155
C.2	Survey Taken by NOLS Students, Cont.	156
C.3	Kernel Density of Age and Proportion Male	157
C.4	Distributions of Individual Composite Abilities, Before and After Course	159

C.5	Pre-Course Individual Ability On Pre-Course Peer Average Ability	160
C.6	Binned Scatter Conditional Randomization Checks for Age	161
C.7	Age on Meanage with Cubic B-Spline	162
C.8	Marginal Effects From Non-Parametric Kernel Regressions with Confidence Inter- vals	165
C.9	Binned Scatter of Pre Course Peer Ability on Post Course Own Ability, Controlling for Own Before Ability	168
C.10	Parallel Analysis	172
C.11	Cumulative Variance	173
C.12	Loading on Components 1 and 2	174

List of Tables

1.1	Main Sample Descriptive Statistics	24
1.2	Alcohol Sales Legalizations Effect on Total Crime at Home Games	26
1.3	Investigating the Influence of Omitted Law Enforcement on Effect of Sales on Crime	29
1.4	Effect of Alcohol Sales on Crime by Type Legal Alcohol Access Age	31
1.5	Effect of Alcohol Sales on Crime Victimization	32
2.1	Pre-Treatment Covariate Balance	73
2.2	Intent to Treat Effect on Meeting Academic Renewal Threshold By Academic El- igibility and Date Cutoff	75
2.3	Local Linear Estimates of Scholarship on First Year Credits and GPA	76
3.1	Composition of Noncognitive Ability Measures	101
3.2	Descriptive Statistics	102
3.3	Own Characteristics Regressed onto Peer Characteristics	103
3.4	OLS of Peer Group on Noncognitive After and Improvement	105
3.5	Heterogeneity in Peer Effects by Course Characteristics	107
3.6	Heterogeneity in Peer Effects by Demographics	108
A.1	Colleges In Sample	125
A.2	Offense Descriptions	126

A.3	Alcohol Sales Legalizations Effect on Total Crime at Home Games, With Exact Date Fixed Effects	128
A.4	Robustness to Model Type	129
A.5	Alcohol Sales Legalizations Effect on Different Crime Types	130
A.6	Alcohol Sales Legalizations Effect on Total Crime at Home Games	131
A.7	Effect of Alcohol Sales on Specific, Related Crime Types	132
A.8	Investigating the Influence of Omitted Attendance and Law Enforcement on Effect of Sales on Crime	133
B.1	Descriptive Statistics by Eligibility	138
B.2	Local Linear Estimates of Scholarship on First Year Credits and GPA	146
B.3	Scholarship on Freshman Credits, Parametric	147
C.1	Descriptive Statistics on Raw Questions	158
C.2	Relationship Between Peer Average Age and Own Age (Within Mean Age Bins)	163
C.3	OLS of Pre-Course Peer Ability on Noncognitive Ability After, By Individual Question	164
C.4	Non-Parametric Kernel Regression of Peer Group on Noncognitive After	166
C.5	Models of Peer Effects	167
C.6	OLS of Peer Group on Instructor Assigned Grades	169

Chapter 1

Not So Fast: Alcohol Sales to College

Football General Seating Sections Reduces Crime

One factor in the growing cost of college is athletic program expenses that continue to increase; many college teams subsidize their athletic programs from general funds, student fees, and government appropriations ([Sackstein, 2019](#)). A threat to profit margins that could further increase athletic expenses in college sports is litigation by student-athletes to receive compensation ([Sanderson & Siegfried, 2015](#)). Attempting to raise additional revenue, many universities have recently removed restrictions on selling alcohol at college football games. From 34 teams allowing alcohol sales in 2015, now at least 78 teams allow sales ([Hayes, 2019](#)). West Virginia University reported \$500,000 dollars in alcohol concession revenue in the first season of sales ([Tracy, 2015](#)) and the University of Texas, earned \$1,100,000 from alcohol sales in its first two home games of 2016 ([Knox, 2016](#)).¹

Allowing alcohol sales at sporting events is controversial. Selling alcohol draws the ire of the

¹Note that raising additional revenue could be consistent with universities being revenue or profit maximizers. Bowen's revenue theory hypothesizes that colleges are revenue-maximizers. ([Bowen, 1980](#)). An alternative model for the growing cost of college is Baumol's Cost Disease ([Archibald & Feldman, 2008](#)).

National Collegiate Athletic Association (NCAA), which bans alcohol sales and advertisements at championship events. The National Institute of Alcoholism and Alcohol Abuse (NIAAA) recommends banning alcohol sales at college sporting events, rating bans as moderately effective for reducing student alcohol consumption (NIAAA, 2015, pg. 19). Alcohol access has harmful consequences for university students and college-aged individuals, reducing academic performance on university exams (Carrell, Hoekstra, & West, 2011a; Lindo, Swensen, & Waddell, 2013) and increasing crime (Carpenter & Dobkin, 2015; Heaton, 2012). In 2010, alcohol abuse cost \$293.4 billion (Sacks, Gonzales, Bouchery, Tomedi, & Brewer, 2015).

Prior work, using crime data from the National Incident-Based Reporting System (NIBRS), finds college football home games are associated with higher amounts of crime (Lindo, Siminski, & Swensen, 2018; Rees & Schnepel, 2009); a leading mechanism proposed is additional alcohol consumption due to gameday festivities. A review on alcohol control and crime argues college football gamedays and the NIBRS dataset are effective tools for understanding the link between alcohol access and crime (Carpenter & Dobkin, 2010). The recent changes to allowing in-stadium alcohol sales enables explicit investigation of whether crime increases on gameday can be attributed to increased alcohol access using the same crime data. That is the research question of this study.

In this paper, differences-in-differences (DD) models estimate alcohol sales to general seating reduces total crime on home game days by approximately 20%. Following Heaton (2012), the control group is non-treated days (days where a home game is not played) in the same city, after sales begin. Factors such as weather, unemployment, changes in student population, or police agency culture affect crime on treated and control days, so it is highly unlikely these alternative determinants of crime bias the estimated effects.

Event-study results find no evidence that the change to sales is an endogenous response to crime. Before sales begin there are no significant effects of home game days on crime. However, after sales begin there is a reduction in total crime that is statistically significant at 95% confidence in several different weeks.

Two factors that may change only on home game days at the same time sales begin, and thus bias the DD coefficient on alcohol access, are law enforcement and other stadium policies. To investigate whether law enforcement changes at the same time as sales begin, the percentage of reports that end in arrests (the clearance rate) is used as the dependent variable in place of crime. The beginning of sales reduced the clearance rate by 8 percentage points on home game days relative to non home game days.

To recover an effect of alcohol access that is not confounded by changes in law enforcement, further analysis leverages a natural experiment in alcohol access created by the minimum legal drinking age (MLDA).² This is the same strategy pursued by [Marcus and Siedler \(2015\)](#) in a differences-in-differences framework. While only those above the MLDA can legally purchase alcohol after the beginning of sales, the change in law enforcement on home game days applies to those above and below the MLDA. The equal change in enforcement is supported by results showing there is an equal decrease in clearance rate for those above and below the MLDA cutoff after sales begin. Some other confounders, like game scheduling differences, apply equally to those above and below the alcohol access age threshold. Other stadium policies, such as additional buses or halftime re-entry, that also may change are difficult to precisely age discriminate meaning they likely apply equally to those above and below the alcohol access age cutoff.

When using age-based variation in legal alcohol access, the negative effect of legal alcohol access on crime at college football games is supported. First, the data is subsampled to only those below the legal alcohol access age, because they do not gain legal access to alcohol after the change to allowing sales. For most crime types, there is no statistically significant effect on this group. For a few crime types, there is a negative effect on crime for this group, but in all the cases where this effect is statistically significant, the negative effect is less negative than the effect for those who are above the legal alcohol access age. Comparing the effect of sales on those above the legal age to those below the legal age suggests that sales reduce home gameday aggravated assault, simple

²This variation is extensively used in age-based regression discontinuity estimates of the effect of alcohol access on crime ([Carpenter & Dobkin, 2015](#); [Lindo, Siminski, & Yerokhin, 2016](#)) and other outcomes ([Carpenter & Dobkin, 2017](#); [Carrell et al., 2011a](#)).

assault, vandalism, rape, and liquor law violations.

Causal studies, outside of college football contexts, overwhelmingly find increased access to alcohol increases crime. Using temporal variation has shown that increased alcohol access increases crime (Heaton, 2012) and reduced alcohol access reduces late-night hospitalizations (Marcus & Siedler, 2015). Similar findings are reported in Biderman, De Mello, and Schneider (2010), Grönqvist and Niknami (2014), Nakaguma and Restrepo (2018), Carpenter and Dobkin (2015), and Carpenter and Dobkin (2017).

Internally valid causal estimates of the effect alcohol access on consumption and crime are environment specific. Carpenter and Dobkin (2015) finds alcohol access increases DUI in California, but Lindo et al. (2016) finds barely exceeding the minimum legal drinking age (MLDA) in New South Wales does not cause increases in DUI. An interesting dimension of using college football alcohol policy changes as variation to measure alcohol access on crime is the change affects a few hours in 1 day as opposed to using weekend-length or age-based alcohol access restrictions.³ Long time periods affected by changes to sales policy make it difficult to understand substitution effects of easing/restricting access to alcohol (Carpenter & Dobkin, 2010).

Studies focused on college football-alcohol sales find mixed results. Due to substitution across time and types of liquor, the three-period, pre-game, game, and post-game, utility maximization model of Boyes and Faith (1993) is unable to predict the expected sign of college football alcohol sales on alcohol consumption.⁴ Some research finds a positive association between college football alcohol sales and crime (Bormann & Stone, 2001; Menaker & Chaney, 2014). Other research finds banning alcohol sales increases the blood alcohol content (BAC) of drivers suspected of driving under the influence (DUI) (Boyes & Faith, 1993). Mixed results call into question the generalizability of prior studies on alcohol access and crime to the college football context and are a dissonance in prior college-football specific research to re-examine with multiple policy changes.

³The length of the intervention is emphasized in non-economics literature, see the review in R. A. Hahn et al. (2010).

⁴The reason the sign is ambiguous in Boyes and Faith (1993) is due to substitution across time and types of alcohol. This substitution across time arises due to bans increasing the cost of in-game alcohol consumption. The substitution across types of alcohol arises from imperfect ban enforcement. It is easier to conceal smaller volumes of higher-proof alcohol which allow an equivalent amount of alcohol consumption as higher volumes of lower-proof alcohol.

1.1 Data

There are 3 primary data sources. Crime data comes from NIBRS, college football game data is scraped from cfbstats.com, and alcohol sales data is manually collected. Each of these is discussed.

College Football College football data comes from www.cfbstats.com.⁵ The data was available back to 2008, so that is the first season in the sample. Post-season games are dropped.

Crime NIBRS provides crime reports by incident.^{6,7} Each incident is counted as one, ignoring the number of offenders per incident.^{8,9} Because the question relates to university policy and college football, the sample is only agencies covering the county or university of a D-1A teams in the NIBRS dataset. Not all college campuses are covered under NIBRS so the sample results in a panel of 44 college football teams, Table A.1 presents these teams, their number of agencies, and how many years each college had sales.^{10,11} While it is true that this is not a random sample of colleges, this data has sufficed in prior research on a very similar topic ([Lindo et al., 2018](#); [Rees & Schnepel, 2009](#)), has been used in other studies of alcohol sales policies ([Heaton, 2012](#)), and is recommended by [Carpenter and Dobkin \(2010\)](#).

⁵This source includes attendance.

⁶For some crimes, designated by the FBI as Type A, there is data by incident including age, time of offense, whether or not alcohol was suspected of being used at time of offense, etc. Type B crimes do not include whether the offender was suspected of using alcohol or time of day that a crime occurred, and generally provide less accompanying information than Type A crimes.

⁷Since the research question is measuring the impact of alcohol sales on criminal behavior, following a suggestion in ([Carpenter & Dobkin, 2010](#)) and dependent variable selection approach of [Rees and Schnepel \(2009\)](#) and [Heaton \(2012\)](#), several different types of crime that alcohol availability could affect are used. Definitions of each crime type is provided in Table A.2.

⁸Most incidents involve 1 offender .

⁹Incidents which record only the date of recording and not the date of the incident occurring are dropped.

¹⁰Other teams that have entered D-IA, FBS, since then are omitted. The omitted teams are Old Dominion, Coastal Carolina, and University of Massachusetts-Amherst. Additionally Eastern Carolina and Miami University (OH) are supposedly covered, but offense data is missing for these two schools. None of these schools changed alcohol policy over the length of the sample. Connecticut is not included because they play many games in Rentschler Field, a professional stadium and are not in the crime data.

¹¹There are 13 agencies that are identified as being connected to a university, but have no offense data in any year are omitted. After dropping these, there is still at least 1 reporting agency for all the teams that change alcohol policy over the length of the sample. That is there is at least 1 reporting agency for 44 teams, resulting in crime data for 79 individual police agencies.

1.1.1 Alcohol Sales

Data for alcohol sales at all NIBRS-covered teams was collected by first searching the college's athletic website. If nothing was found, websites of local news outlets were referenced. Seventeen agencies across 8 teams report crime in the same city as a college that changed to allowing alcohol sales to general seating.¹² Figure 1.1 shows the spatial distribution of colleges covered by NIBRS and color codes for which years teams changed to allowing sales, if they ever did. Teams that allowed sales or a designated beer garden for general admission fans are counted as starting alcohol sales.

Dates and Imbalance From Variation in Treatment Timing The dates that are used are the length of the available college football data, 2008-2015. The data is daily, dropping May-July like Lindo et al. (2018), because days in these months are not during the typical college school year.¹³ For teams that change in 2009, there is only 1 season before the change in the data. For teams that change in 2014, there are 6 seasons before the change. There are up to 6 seasons before and up to 7 seasons after, but only 1 pre-season observed for all changing teams and only 2 post-seasons observed for all changing teams. Data is sometimes setup using event weeks as the panel time variable to address imbalance of treatment in the balanced panel.

1.1.2 Descriptive Statistics

Combining these data sources results in balanced panel of agency-by-day level observations including crime counts, college football game information, and the beginning of alcohol sales. Table 1.1 shows descriptive statistics. Panel A shows the aggregate sample. 2 percent of the sample are home games and 14 percent of days are after sales have been legalized at games. Game And/Or Alcohol-Related Crimes stands for crimes associated with college football or alcohol and include

¹²Data on changes in alcohol sales have been cross-checked with one other researcher's list.

¹³Lindo et al. (2018) uses more years- 1992-2012, which may not be obviously beneficial for this research question, since the research design leverages a policy change.

assaults, rape, vandalism, DUI, disorderly conduct, liquor law violations, and drunkenness.¹⁴

Panels B-G subsample to only teams that changed their policy and split the data by pre-period/post-period and home, away, and non-game days. Comparing Panels B and C shows mean game-related crimes increased in the post period, by about 3 crimes, and total crime increased, by about 11 crimes, for home games. In contrast, there was a decline in median crime, total and game-alcohol related, after changing to allowing sales on home game days. The reduction in median total crime is from 16 to 7, a 56% reduction, and the reduction in median game-alcohol related crime is a 67% reduction on home game days. The difference in how crime changes after allowing sales in mean and median is likely due to outliers affecting the mean more severely.

Panels D and E show there was a reduction in median crime on away games also, however, the reduction is smaller in magnitude than the reduction on home game days. Particularly, there was a 50% reduction in median game-alcohol related crimes for away games after the change (4 before, Panel D; 2 after, Panel E). After changing to allowing sales, there is also a 50% reduction in median game-alcohol related on non-gamedays. The magnitude of the reduction in crime is the exact same on away games as it is on non-gamedays. After the change to allowing sales, the reduction in median crime on away and non-gamedays is smaller than the reduction in median crime on home gamedays.

1.2 Method

Differences-in-differences estimates how stadium alcohol sales affect crime. The equation is

$$Crime_{pcd} = \exp(\beta_1 Home_{cd} + \beta_2 Sales_{cs} + \beta_3 Home * Sales_{cds} + \mu_P + \lambda_T + \phi_O), \quad (1.1)$$

in which p represents police agency, c represents colleges, s represents college football seasons, d represents days, T represents time, and O represents day of week order. Sales is a dummy variable

¹⁴The selection of crime types is based on these types being used in [Rees and Schnepel \(2009\)](#) or [Lindo et al. \(2018\)](#).

equalling 1 if it is August or after of the year the college football team began selling alcohol.¹⁵ Home is a dummy variable equal to 1 if a college football game was played within the jurisdiction of a reporting agency.¹⁶

The standard errors in the preferred specification are clustered at the city level to adjust for the possibility that errors are correlated over time within groups. The dependent variable is a count variable, so Poisson models primarily are used.¹⁷ At first, the dependent variable is total crime, then robustness checks subset to associated and individual crime types.

Agency fixed effects, μ_P , control for time-invariant characteristics of agencies over the length of the sample, for example climate or region. One could argue some things such as agency culture are relatively constant over the 8 year sample period. Time fixed effects, λ_T , control for aggregate variation in incidents that could potentially be related to game scheduling trends over time.

Day of week indicators, ϕ_O , are important, because college gamedays only occur on certain days, usually Saturday. Although most college games occur on Saturdays, around half of a teams schedule each season is played in another teams' stadium outside of the home agencies' jurisdiction which leaves some Saturdays with no home game during the fall semester. Day of week is also correlated with crime, specifically because weekdays are typically associated with greater amounts of crime. Additionally, if universities adjust their schedule at the same time as the ban, for instance by moving Saturday games to Thursday, then including day of week adjusts for this observable difference.

Although not shown in Equation 1.1, the model also includes indicator variables for holidays like Lindo et al. (2018). The holiday indicators control for changes in activities that are associated with holidays.¹⁸ In addition to reasons in Lindo et al. (2018), colleges could be less likely to

¹⁵Sales beginning is the beginning of the post period.

¹⁶Equation 1.1 applies the (Sunday) alcohol sales legalization main equation from Heaton (2012) to the college football home game crime question estimated in Rees and Schnepel (2009) and Lindo et al. (2018).

¹⁷Poisson estimation is favored in Heaton (2012) and Lindo et al. (2018) and has relatively favorable theoretical results compared to negative binomial models. (Allison & Waterman, 2002; Cameron & Trivedi, 2005; Wooldridge, 1999). In particular, Cameron and Trivedi (2005) shows that cluster-robust standard errors in Poisson models correct for over-dispersion which defeats the main purpose of including an overdispersion parameter (i.e. using negative binomial models.) Results are not qualitatively different when using of negative binomial or OLS.

¹⁸The holiday controls included are: Columbus Day, Halloween, Labor Day, Thanksgiving, and Veterans' Day.

schedule games on holidays associated with drinking after alcohol sales are allowed.¹⁹

Theory of Alcohol Sales and Alcohol Consumption A useful theoretical model for considering stadium alcohol access and alcohol consumption is a three-period utility maximization problem. The 3 periods are pre-game, game, and post-game (Boyes & Faith, 1993). In the model, the effect legal alcohol sales in college stadiums on alcohol consumption are ambiguous due to intertemporal substitution of consumption (Boyes & Faith, 1993). Other relevant considerations include the price of alcohol, the degree of imperfect enforcement, and the substitution between low and high proof liquors due to high proof being easier to conceal.

Mechanisms from Alcohol Consumption to Crime Goldstein (1985) describes 3 mechanisms from alcohol consumption to crime. The first mechanism is a direct pharmacological affect, where increased blood alcohol content (BAC) reduces the cognitive function of individuals which can lead to poor decisions. The second mechanism is the excuse effect, where an individual believes being intoxicated will lead to less harsh sentencing if convicted. The third mechanism is the social interaction effect, which is essentially that alcohol use is often positively correlated with social activity, so increasing alcohol use also increases the number of social interactions one has which independently could lead to additional crime.²⁰

1.2.1 Selecting a Control Group

Equation 1.1 estimates a treatment effect of alcohol sales by comparing home games to non-home games in the same agency, before and after the change. The approach uses home game days, where a college played a game in the jurisdiction of an agency in the same city as a college in a given day, of teams that changed to allowing alcohol sales as a treated group. This design takes advantage of

¹⁹Lindo et al. (2018) notes a “direct” bias from the association between holidays and the days which games are played on. Another way is through an “indirect” bias through the day of week fixed effects, because certain holidays always fall on certain days of the week.

²⁰It is empirically challenging to differentiate which channel alcohol use affects outcomes through (Carpenter & Dobkin, 2010).

days where no home game was played, meaning no college football game at all or the team was playing but not in the same city or stadium, as a control group.²¹

In addition to the home/treatment day versus non-treatment days used by previous papers (Heaton, 2012; Lindo et al., 2018; Rees & Schnepel, 2009), this strategy is appropriate because teams that do not change may be systematically different than teams that do change.²² For example, teams that change are mostly geographically similar by being from the midwest (see Figure 1.1) and they are not from the most competitive conference (the SEC) due to the SEC banning alcohol sales over the sample period. These reasons make non-changing teams a less attractive counterfactual, using non-home game days is not vulnerable to these criticisms. Conditional on away games not changing after sales, away games could make an attractive control group, because it is on a similar day at the same time of the year.

1.2.2 Identification

The identification assumption is that in the absence of the change to alcohol sales, the home game-days and non-home game days would not diverge from their common crime trends. The high variance, relative to the mean, of crime data makes this difficult to observe in typical DD parallel trends. It is reasonable to believe that changing to sales was done with the main purpose of raising additional revenue, not in response to home game crime trends.

Equation 1.1 omits several variables which may be related to crime. Some variables correlated with crime are unemployment and number of law enforcement officers deployed. The reason unemployment and number of law enforcement officers are omitted is because data does not exist at the agency level with daily frequency.

The DD coefficient, β_3 from Equation 1.1 is unlikely to be biased by omitting daily unemployment. For excluding unemployment to bias the parsimonious DD coefficient, unemployment

²¹In contrast to typical DD applications, one can drop all teams that did not change and still estimate Equation 1.1 due to the presence of the untreated non-home games. Leveraging non-treated days of treated teams means non-changing teams can be completely removed and there is still identifying variation.

²²If changing and non-changing teams are dramatically different, then teams that do not change would be an undesirable counterfactual, because it would not be clear that the policy change is driving the estimated effect.

would have to change at the same time as the policy. Furthermore, because the control group is non home game days, the change in unemployment would have to affect home game days differently from all other days. Law enforcement is more likely to bias the parsimonious DD coefficient, because it conceivably changes only on home gamedays.²³

1.3 Results

First, a general difference-in-differences model is estimated and robustness checks performed. Next, event study models are estimated to investigate time-varying treatment and endogenous policy adoption. Next, the same differences in differences model is estimated with clearance rates as the dependent variable to investigate changes in enforcement as a potential confounder. Following confirmation that clearance rates change, additional variation in alcohol access, due to the minimum legal drinking age, is shown to address changes in enforcement. Finally, DD models are estimated separately for those above the MLDA and below it.

1.3.1 Home and Away Game Premiums in Pre and Post Periods

[Figure 1.2](#) presents stylized versions of parallel trends. To mirror the estimation strategy which does not compare home games of treated teams to home games of non-treated teams, it incorporates crime on non-treated days in the post-period for treated teams. Only teams whose alcohol sales policy changed are included.²⁴

One particular day, away games, could be a good control group. Away games could be good, because the sales policy does not affect days when there is no game played in that location, but away games are played at the same time of the year on the same day of the week. To leverage away games, [Figure 1.2](#) keeps gamedays (home and away) and non-game days which are 2 days to either side of a home or away game.

²³Enforcement being time-varying, correlated, omitted variable leads to the necessity of leveraging age-based variation after the parsimonious DD.

²⁴This approach closely follows the scatterplot graphical evidence presented in [Heaton \(2012\)](#).

Next, the non away game day crime median (from 2 days on either side of the away game) is subtracted from the away game crime median. Also, the non home game day crime median (from 2 days on either side of the home game) is subtracted from the home game crime average. Call this variable the gameday crime premium. For example,

$$\text{Home Gameday Crime Premium}_w = \text{Home Game } \bar{C}_w - \text{NonHome Game } \bar{C}_w \quad (1.2)$$

, where the subscript w stands for event week and \bar{C} stands for “crime-median”.²⁵ The y-axis plots the extra crimes that occurred on a gameday compared to a non gameday.

The x-axis is centered around the number of games away from the first season/game of alcohol sales. Using seasons away would not produce a useful comparison, because there is only 1 pre-period and 2 post-periods for all changing teams. The black dashed vertical line at 0 represents the change to allowing sales.²⁶

Figure 1.2 shows the home gameday crime premium decreases from 12.1 to 5.7 after stadium sales begin. The average away gameday crime premium only slightly changes from 2.3 to 1.85 after sales are introduced. The home gameday crime premium falls and the away gameday crime premium does not, suggesting allowing alcohol sales is not correlated with an overall decrease in weekend crime.²⁷

1.3.2 Difference in Differences Results

Table 1.2, Panel A presents the main results of estimating Equation 1.1 using models increasingly saturated with fixed effects. In each column, the estimated effect of home games is positive and the estimated effect of allowing sales at home games is to reduce the total amount of crime on home games relative to non-home games. Both the coefficients on home and home X sales are statistically different from 0 at greater than 95% confidence in all specifications, although they

²⁵Median is chosen over mean to reduce the impact of outliers on the results. Table 1.1 shows the median and mean are different suggesting outliers influence the mean as a measure of central tendency.

²⁶Going from a shaded to non-shaded background (and vice-versa) in Figure 1.2 represents beginning a new season.

²⁷Figure A.2 shows crime on home games compared to away games decreases after sales begin.

have opposite signs. When the day of week and holiday indicators are added (column 4), home and home X sales coefficients have the exact same magnitude in opposite directions. Home games are estimated to increase total crime by approximately 20% which is completely eliminated after the sale of alcohol.²⁸ The mean of daily crime is 29.35, making this is equal to about 5.57 crimes.²⁹

Table 1.2, Panel B presents results of estimating Equation 1.1 using models that also include a dummy variable for away game and an interaction between away and sales. If the decrease is not caused by weekend crime trends, then AwayXSales should be 0 and not statistically significant. The inclusion of away and awayXsales does not change the coefficient on HomeXSales at all. AwayXSales is statistically significant in Panel B, columns 2 and 3, but once day of week fixed effects are included in column 4, the coefficient goes to 0 and is no longer statistically significant. Comparing away game crime, after sales, to crime on the same day of the week, without a home a game, finds no effect of changing to sales on away crime.³⁰

The negative sign on the DD coefficient suggests allowing alcohol sales reduces overall crime, which departs from most of the general results on alcohol access and crime (Carpenter & Dobkin, 2015; Heaton, 2012), but is similar to Boyes and Faith (1993). Four competing hypotheses for the un-anticipated, negative sign exist: substitution across time and/or types of liquor, time-varying omitted variables which bias the coefficient estimate, large increases in crime in the pre-period which the policy was designed to address, or poor sample/dependent variable choice. Further analyses investigate the latter three, seeking to rule them out leaving the first reason as the only explanation.

Robustness Checks First, Table A.5 separates total crime into each disaggregated crime type. There are no positive effects for any crime, the effects are concentrated in crimes associated with binge drinking at college football including assault types, disorderly conduct, and liquor law violations, and there are no effects on grand theft auto, embezzlement, or peeping tom. Panel A of

²⁸The coefficient magnitude is calculated: $(e^{0.19} - 1) * 100 \approx 20$ percent.

²⁹Table A.4 shows that the DD coefficient is negative and statistically significant also when using OLS or Negative Binomial models. This suggests the results are also not driven by type of regression model used. Also, OLS estimates a very similarly sized DD coefficient, 5.82 crimes.

³⁰Table A.3 estimates the same equations, using exact date FE's, and the results are exactly the same.

Table A.6 aggregates these crime types. The coefficient is smaller when liquor law violations are dropped out, which could be expected given the policy makes liquor legal. The DD coefficient gets larger when using only the crimes that are associated with games in columns 2 and 3. Columns 4 and 5 narrow down the crimes to only those where alcohol is involved, column 5 additionally subsamples to within 3 hours of game time, and finds similar results. There are similar effects when only gamedays are included in the estimation sample and there are effects for both university and city/county agencies.³¹ Finally, Table A.7 finds that the crime reductions happen in locations that are associated with college football and/or alcohol consumption.

1.3.3 Inference Robust to Few (Treated) Clusters

The standard errors are clustered by college, which accounts for within college correlation of the standard errors; but there are two reasons to be suspicious that the DD estimates have misleadingly narrow standard errors. The issues are there are only 44 colleges by which to cluster and coefficient is estimated from only 9 teams changing. These issues mean confidence intervals could be inappropriately narrow (Cameron & Miller, 2015). To address the potentially narrow standard errors, a permutation test is used which is less vulnerable to issues erroneous inference resulting from few clusters (Fisher, 1935).³² This procedure is based on Fisher's exact p-value.

For the procedure, any number of teams are randomly assigned a year of beginning sales.³³ The test statistics from each of the 1,000 permutations are displayed in a histogram in Figure 1.3. The actual test statistic of the coefficient on HomeXSales is outside the 1st percentile (The vertical green line with dots) of the placebo distribution of test statistics. The number of times one of the z-statistics on β_3 from Equation 2 from falsely randomly assigning treatment was less than the actual z-statistic of home X sales coefficient from Equation 1.1 with realistically assigned treatment was

³¹University agencies are used by both Lindo et al. (2018) and Rees and Schnepel (2009) and there is no reason to believe that any potential double counting changes at the same time as the sales begin. Thus any concerns about measurement error are only classical measurement error leading to attenuation bias.

³²This is a similar approach to deal with narrow confidence intervals from few clusters as Cunningham and Shah (2018).

³³Specifically, all teams are randomly assigned season/year for beginning sales. The date is August 1, random year. Whether random assignment of any number of teams is appropriate depends on the true data generating process.

only 13/1000 repetitions, which implies an exact Fisher p-value of $13/1000 = .013$. No evidence is found that statistical significance is solely due to few overall clusters or few treated clusters.

1.3.4 Event-Study

An event-study framework is adopted to provide evidence related to endogenous adoption of alcohol sales. The objective is comparing home game crime changes in the post-period to home game crime in the pre-period. The equation used is

$$Home\ Crime_{pw} - Away\ Crime_{pw} = \sum_{w=-30}^{-2} \delta_w Week_w + \sum_{w=0}^{34} \delta_w Week_w + \mu_p + e_{pw}, \quad (1.3)$$

in which w stands for event week and all other subscripts are similar. Only gamedays and changing teams are included in the estimation sample. The $Week$'s are binary variables equal to 1 if it is a specific event week and δ_w is the associated coefficient estimate.³⁴ As is standard, the event week of -1 games from the allowance of sales is omitted so coefficients compare crime in other event weeks to this period. The dependent variable $Crime_{act}$ is adjusted for away games by subtracting the away game crime to leverage non-treated days in treated groups.^{35,36} It is preferred to include fixed effects for agencies, because it is possible for teams to drop out of the sample if the agency stops reporting data.³⁷

The event-study is non-standard, because of the data and the empirical strategy. With the data it is impossible to go by seasons, instead of games, away from sales, because there is 1 season of pre-sales data for all teams in the sample. Next, the dependent variable is home-away, because the

³⁴As discussed in Section 1.1.1, the balanced panel and variation in treatment timing lead to uneven temporal coverage. All lags and leads are included, but there is only 1 pre-season and 2 post-seasons where all colleges are included, so these coefficients are presented. A different approach is pooling all years outside of these years. Some discussion of pros and cons of pooling data in event studies can be found in [McCrary \(2007\)](#).

³⁵This also necessitates using OLS, because there can be negatively-valued observations of the dependent variable.

³⁶Furthermore, there are not many observations, so using away game crime increases the amount of data, potentially offering efficiency gains. The drawback is a trend in away game crime could bias the estimate, this is not concerning given the insignificant effect the change has on away game crime.

³⁷Only teams that change to allowing stadium sales are kept, which facilitates construction of the appropriate comparison group, reduces variables in the model, and eases interpretation. There are only a few teams who change the policy meaning that dropping teams leaves less statistical power in the model. Due to lack of power, month, year, and sales variables are not included. Including fixed effects leaves too little variation to yield any precise results.

main methods compares home to non-home days, instead of treated team home to non-treated team home. So, it is necessary to also adjust the event-study approach to use the same counterfactual. That is why the dependent variable is home-away. Away games will often be played on the same weekend days, so it makes sense to do this.

Figure 1.4 presents coefficients from each period with heteroskedasticity-robust 95 percent confidence intervals.³⁸ There are no significant effects ($p < 0.05$) of sales on crime before sales are allowed, whereas a pre-trend would be suggestive the policy is endogenous. The joint F-statistic that all pre-period coefficients are 0 is 1.14 ($p = 0.339$), meaning the null hypothesis of no pre-trends cannot be rejected. The event study approach finds no evidence endogenous policy adoption is responsible for the negative effect of alcohol sales on crime.

1.3.5 Investigating Omitted Variable Bias

While the control group of non-home game days deals with many potential omitted variables, there are some omitted variables that may change only on home game days, after sales begin, that are correlated with crime. Two such variables on which some kind of data can be collected are game attendance and law enforcement strategy. Each of these are investigated.

Attendance Attendance could change, since more spectators may attend in response to alcohol being made available. Additional attendance would lead to increased likelihood of crime through a social interaction effect instead of a direct pharmacological effect (Goldstein, 1985). Table A.8 finds that including attendance in the regressions do not meaningfully affect the DD coefficient.³⁹

Law Enforcement

Table 1.3 investigates whether law enforcement strategy plausibly changes at the time sales begin. Law enforcement strategy is almost certainly related to the amount of recorded crime. Also, there

³⁸City-clustered standard errors produce similar results, despite there are being only 8 city clusters.

³⁹It is challenging to include attendance, because it is only observed on gamedays, but the comparison group is non-home gamedays.

are several examples of statements made about law enforcement changing, on home gamedays only, after the change to sales (e.g. ([Penn Live, 2011](#))). It is less straightforward to sign the bias for law enforcement, because law enforcement could deploy more or less officers after the change and more officers could lead to more crime ([Levitt, 1995](#)) or less crime ([MacDonald, Klick, & Grunwald, 2016](#)).

Controlling for law enforcement strategy is also not straightforward, because the daily count of deployed officers is not observed.⁴⁰ Prior studies using NIBRS use clearance rates as the best daily, observable measure of law enforcement strategy ([Heaton, 2012; Lindo et al., 2018](#)), because [Mas \(2006\)](#) finds that clearance rates are positively correlated with deployment numbers.⁴¹ Clearance rate is used as the dependent variable, in prior studies, to see if the policy affects the likelihood of being arrested, which would be consistent with changes in enforcement.⁴² If the DD coefficient does not change the arrest rate on home games after the change to sales, compared to non-home games, the data is not consistent with estimated effects being driven by changes in enforcement.

[Table 1.3](#) estimates the effect of changes to sales on home game day clearance rates.^{43,44} Panel C shows that, regardless of how one treats the missing values (where there are no reports), the effect of the DD coefficient is to reduce the clearance rate on home games compared to non-home games. This is a problem, because a reduction in clearance rates implies lower enforcement intensity, which could be responsible for reductions in crime.⁴⁵ Panel D investigates the same outcome for away games and finds that this reduced enforcement is not present for away games.

⁴⁰Strictness of enforcement, areas, or times of deployment also likely matter, but are less likely to be observed.

⁴¹Clearance rate is the number of arrests per report ($\frac{\text{number of arrests}}{\text{number of reports}}$). More deployed officers lead to higher clearance rates, because more officers means officers can respond faster leading to the likelihood that a report generates an arrest.

⁴²Prior studies ([Heaton, 2012; Lindo et al., 2018](#)) interpret the lack of an effect of the policies under consideration on the clearance rate as evidence that omitting law enforcement variables does not bias the coefficients of interest.

⁴³The panels no longer use Poisson models, because the dependent variable, clearance rate, is not a discrete variable.

⁴⁴Results are not different when using models that explicitly account for the dependent variable being bounded by 0 below and 1 above, such as fractional logit or zero-inflated binomial models. These results are available upon request.

⁴⁵It also suggests that the null results of [Lindo et al. \(2018\)](#) and [Heaton \(2012\)](#) are meaningful, because it rejects that there is no change in enforcement when there is anecdotal evidence that this may be the case.

1.4 Using Age-Based Variation to Address Omitted Variables

Table 1.3 suggests omitting enforcement may bias DD estimates and be responsible for the unexpected negative sign. To address likely omitted variable bias from time-variant, law enforcement, Section 1.4 uses age-based variation in legal access to in-stadium alcohol, that also varies over time (see Figure A.1). This change in policy only affecting a specific age is similar to Lemieux and Milligan (2008), where individuals of a certain age gain access to benefits after a point in time and is the same strategy used by Marcus and Siedler (2015). Those exceeding the minimum legal drinking age were allowed to purchase alcohol and exposed to the same time-varying confounders as those who did not gain access. Age-based variation addresses bias from omitting enforcement if enforcement changes equally affect both underage and legally aged individuals.⁴⁶

1.4.1 Clearance Rate at Legal Drinking Age Cutoff

First, whether clearance rates are continuous at the minimum legal drinking age cutoff is examined in the spirit of a regression discontinuity design. Any discontinuous changes in the conditional mean of the clearance rate at the minimum age cutoff age would be a violation of the necessary condition for enforcement to change smoothly at the cutoff, which is the argument for leveraging age based variation. This does not claim that the treatment effect of turning 21 on clearance rate is non-parametrically identified, since there are no observations infinitesimally close to the cutoff (Lee & Card, 2008); however, the quantity can be parametrically identified and many studies use discrete running variables (Figlio & Özek, 2020; Scott-Clayton, 2011), like age is here.⁴⁷

Figure 1.5 tests the continuity of the home game clearance rate at the age cutoff, before and after

⁴⁶Heaton (2012) uses counties that do not change as a triple difference. In contrast with Heaton (2012), at least some unobservable variables which could bias the key DD estimand, due their correlation with the treatment (sales) and outcome (crime) and subsequent omission from the parsimonious specification, are not well controlled for by using unaffected counties/teams. Using non-treated counties as a triple difference theoretically addresses omitted variables such as TV advertisements in Heaton (2012) because it is entirely plausible that TV advertisements are aired for whole regions. The same strategy does not work for stadium alcohol policy, because omitted variables like enforcement only change for treated college teams on home games.

⁴⁷The test may have low power due to not having many observations and not having observations closer than “1 year” away from the cutoff on either side. This is due to NIBRS only recording age of offender to the year.

sales are allowed. There is not a visible discontinuous increase in clearance rate at the age cutoff. If anything the higher clearance rate of underage individuals converges downwards towards legally aged clearance rates after sales, which suggests omitting enforcement will reduce the likelihood of finding less crimes for overage people (i.e. being consistent with the results of Section 1.3).⁴⁸ No different effects of sales on clearance rates at the cutoff supports leveraging age-based variation to adjust for omitted enforcement bias.

1.4.2 Effect of Alcohol Access on Crime, Using Variation From Age

To leverage age-based variation, two separate, related regression approaches are used. First, the sample is split into those who are above the legal age and those below it. The subsample is done, because other policy changes could also affect crime from those unaffected directly by sales. Second, triple differences uses legal age as a third difference.⁴⁹ The underage group is 18-20 year olds and the legal age group are 21-23 year olds.

Table 1.4 shows the results that use age as variation in alcohol access. In Panel A, column 1, the DD coefficient for those above the minimum legal drinking age is negative and statistically significant for aggravated assault and column 2, for those below the minimum legal drinking age, is not statistically significant. Column 3 shows the triple difference is the difference between the difference in difference coefficient. It is interpreted that the change to sales reduced crime for those above the minimum legal drinking age compared to those below the minimum legal drinking age for home game days by about 17%.

Column 4 shows that there was a 37% (.32) reduction in simple assault for the overage and a 22% reduction for underage. That makes the treatment effect a 10% reduction in simple assault. It is likely that simple assault was reduced for those underage due to other policies that happened at the same time. However, those other policies affect both under and overage individuals, so the difference between over and underage individuals is that overage individuals gain access to legal

⁴⁸Figures A.3a and A.3b show that there is not a noticeable change at the age cutoff of at least 1 report being filed on gameday.

⁴⁹Also fixed effects RD estimator, mimicking Pettersson-Lidbom (2012), is used.

alcohol. Columns 7-9 of Panel A show there is no effect for any group on intimidation.

Panel B columns 1-3 show that there is a reduction in disorderly conduct for over and underage individuals, so there is no treatment effect on disorderly conduct. Panel B shows that there was a 43% (.36) reduction in vandalism for underage and a 14% reduction for overage, suggesting a treatment effect 22% reduction in vandalism due to alcohol sales.

Panel B, column 7 shows that sales reduce rape for those who are above the legal age on gamedays. Panel B, column 8 shows that sales increase rape for those who are under the legal drinking age. Both of these coefficients are somewhat large, the coefficient in column 8 is a 61% (.48) magnitude and the one in column 7 is a 43% (.36) magnitude. This implies that allowing alcohol sales reduced rape by 127%, which is large and rape is a rare outcome. This is suggestive, but should be taken with a grain of salt. It does suggest that alcohol access affects rape, but counter-intuitive because legalizing alcohol sales reduces rape. It suggests that prohibited alcohol leading to a period without legal access might actually increase the sexual assault victimization found in [\(Lindo et al., 2018\)](#).

Panel C shows DUI in columns 1, 2, and 3 and there is no statistically significant effect, because underage and legally aged have an approximately equal decrease. This could be due to additional bus routes to and from game, that apply to both ages. Panel C finds a very large reduction in liquor law violations for both ages, but for overage it is much larger than underage. This is because it is no longer illegal for them to buy in the stadium. There is also no statistical difference among the over and underage for getting charged with drunkenness after the change to sales. Panel D presents several falsification checks with crimes that are less likely to be affected by alcohol sales. Most find no effects. Overall sales reduce crime more amongst overage than underage individuals for aggravated assault, simple assault, vandalism, rape, and liquor law violations.

1.4.3 Effect of Alcohol Access on Victimization, Using Variation From Age

Table 1.5 repeats the analysis from Table 1.4, after replacing crime commission with crime victimization.⁵⁰ Panel A shows that total victimization is reduced by 34.9% (.3) for legally aged individuals and reduced by 21% for those who do not gain legal access to inside the stadium alcohol. The resulting treatment effect is a 10% reduction in total victimization.

Panel B shows that legally aged individuals have a greater reduction (44%) in simple assault victimization than those who are underage (23%). This suggests that alcohol access reduced simple assault victimization by 14%, an effect that is statistically significant at 95% confidence. There is a similar effect shown for aggravated assault in Panel C. Panel D and E find no effects on intimidation or rape.

1.5 Conclusion

With aims of higher revenues, colleges have begun to legalize alcohol sales at college football games en masse. While increased alcohol access has been shown to be responsible for crime in many settings and lower academic performance of college students, the empirical evidence of the effect on crime at college football games is mixed and the theory is ambiguous on how sales affect consumption. Using multiple changes to allowing sales at college football stadiums, this research investigates whether alcohol sales affects crime.

The results suggest that alcohol sales reduces home game day crime by about 20%, compared to non-home game day crimes. The event study finds no evidence that the policy is endogenously adopted. While law enforcement strategy (proxied by clearance rates) do change on home game days, the clearance rate does not change discontinuously across the minimum legal drinking age threshold, meaning that law enforcement changes do not bias the effects that are calculated by comparing legally aged to underage individuals. Calculating treatment effects this way, the change to alcohol sales and easier gametime alcohol access are still estimated to reduce total crime.

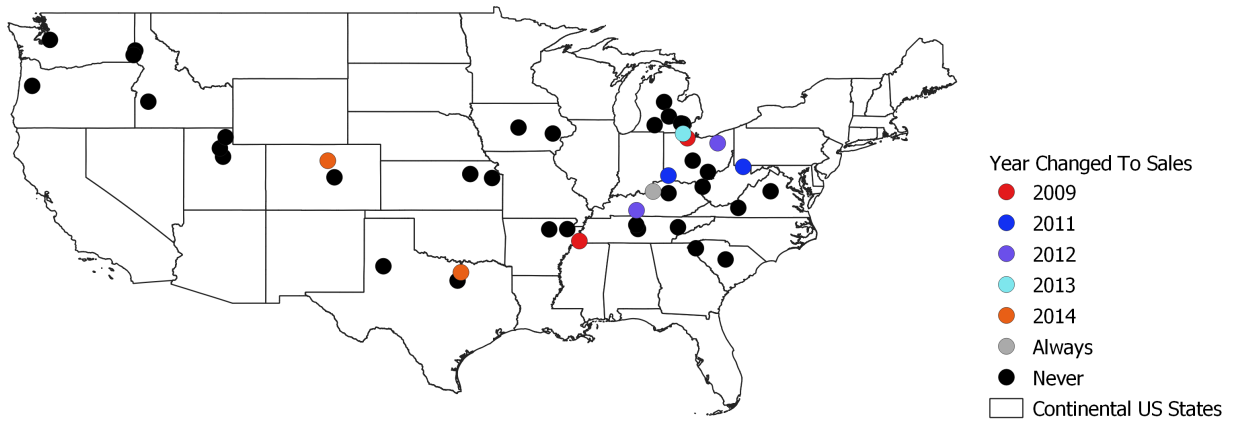
⁵⁰Only Type A crimes report victim characteristics.

The main limitations are the available data. If exact birthdate were observed, then it could enable the use of regression discontinuity methods, such as difference in discontinuities (Grembi, Nannicini, & Troiano, 2016). There are only limited teams in NIBRS, treatment effects might be different for teams outside of NIBRS. The ideal natural experiment would flip a coin for alcohol sales, but this is unlikely to ever happen, making this the best natural experiment in college football stadium alcohol sales that is available.

There are interesting policy implications. It suggests the allowance of alcohol sales at college football stadiums is a win-win. Universities increase concession revenue, and possibly ticket revenue, and the city experiences lower negative externalities from crime. Nonetheless, there are strongly held beliefs on alcohol access and its effect on consumption and crime making it potentially difficult to implement these changes widely.

There are several important directions for future research. First, are there other interventions which affect alcohol access for only a short period? If so, these should be used to gain a greater understanding of intertemporal substitution of alcohol consumption. Furthermore, it would be interesting to observe the effects of stadium alcohol sales on alcohol sales from non-stadium sources. This might provide clues to how consumption outside the stadium changes. Finally, are there similar effects in other samples and for other sources of crime data?

Figure 1.1: Years Teams in Sample Changed to Sales



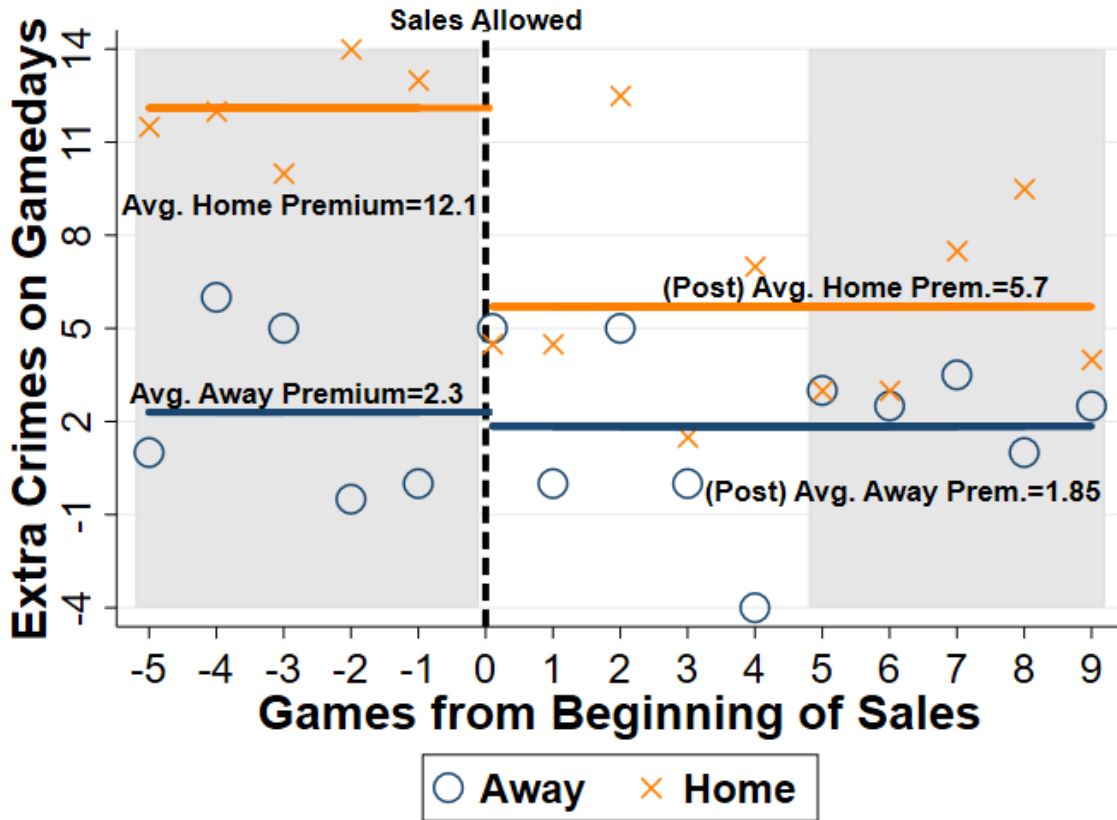
Notes: Only includes D-1A colleges represented in National Incident-Based Reporting System (NIBRS) Data.

Table 1.1: Main Sample Descriptive Statistics

Panel A: Aggregate						
	Mean	Median	Std Dev	Min	Max	Count
Total Crime	29.35	5	66.30	0	613	162898
Game And/Or Alcohol-Related Crimes	10.85	2	25.42	0	254	162898
Home	0.02	0	0.15	0	1	162898
Sales	0.14	0	0.35	0	1	162898
Panel B: Home, Pre-Sales						
	Mean	Median	Std Dev	Min	Max	Count
Total Crime	33.09	16	63.50	0	429	313
Game And/Or Alcohol-Related Crimes	14.95	9	25.76	0	185	313
Panel C: Home, Post-Sales						
	Mean	Median	Std Dev	Min	Max	Count
Total Crime	44.51	7	92.24	0	424	548
Game And/Or Alcohol-Related Crimes	18.61	3	39.32	0	205	548
Panel D: Away, Pre-Sales						
	Mean	Median	Std Dev	Min	Max	Count
Total Crime	28.02	9	59.62	0	453	297
Game And/Or Alcohol-Related Crimes	11.74	4	23.49	0	186	297
Panel E: Away, Post-Sales						
	Mean	Median	Std Dev	Min	Max	Count
Total Crime	41.47	5	90.09	0	420	534
Game And/Or Alcohol-Related Crimes	17.23	2	39.41	0	203	534
Panel F: Non-Game, Pre-Sales						
	Mean	Median	Std Dev	Min	Max	Count
Total Crime	25.18	5	58.64	0	613	13559
Game And/Or Alcohol-Related Crimes	9.53	2	22.62	0	214	13559
Panel G: Non-Game, Post-Sales						
	Mean	Median	Std Dev	Min	Max	Count
Total Crime	40.79	4	89.94	0	470	21758
Game And/Or Alcohol-Related Crimes	15.91	1	37.18	0	221	21758

Note: Panel A includes all game and non-gameday observations, including teams that do not change policy. Panels B-F only include teams that change their policy over the length of the sample. Home is a dummy variable = 1 if there is home game on that day. Sales is a dummy variable = 1 if it is a team that changed to sales and in the post-period. Total crime is all crime types. Game And/Or Alcohol-Related Crimes are: rape, simple assault, aggravated assault, intimidation, vandalism, driving under the influence (DUI), disorderly conduct, liquor law violations, and drunkenness.

Figure 1.2: Gameday Crime Premiums for Home Game Versus Away Game Weeks



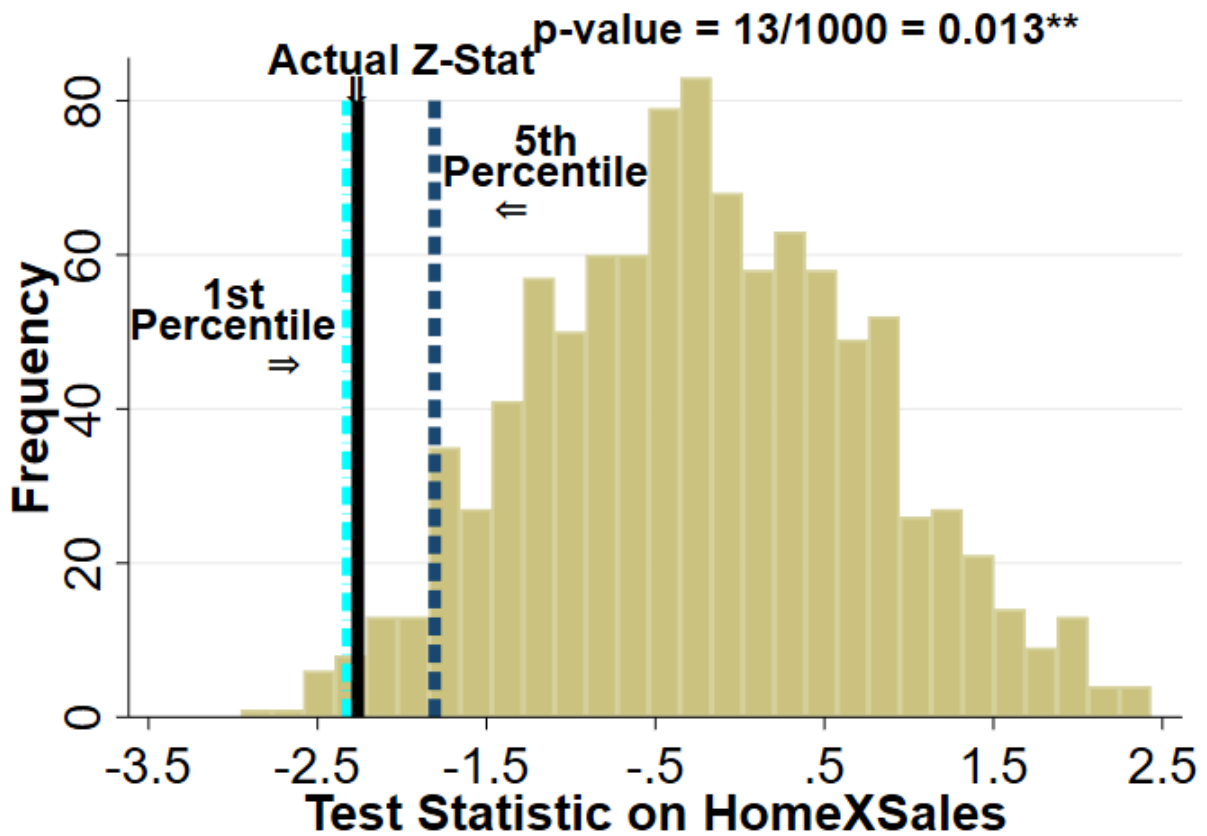
Note: N = 2,100 agency-date level observations. The Home and Away series are the difference in median gameday crime between the day of the game and the two days after/before. Black dashed vertical line represents beginning of alcohol sales. Changing background shade indicates beginning of new season. 3 seasons of data reported, because it is balanced for all teams/agencies. $HomePremium_{ew} = (HomeGame\bar{C}_{ew} - NonHomeGame\bar{C}_{ew})$

Table 1.2: Alcohol Sales Legalizations Effect on Total Crime at Home Games

Panel A: Without Away and Away Interaction					
	(1)	(2)	(3)	(4)	(5)
Home	0.31***	0.28***	0.24***	0.19***	0.19***
	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)
HomeXSales	-0.21**	-0.19**	-0.19**	-0.19**	-0.19**
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Observations	162898	162898	162898	162898	162898
Number of Agencies	79	79	79	79	79
Agency FE	-	X	X	X	X
Years	-	-	X	X	X
Months	-	-	X	X	X
Week of Yr	-	-	-	-	X
Day of Week	-	-	-	X	X
Holidays	-	-	-	X	X
Panel B: With Away and Away Interaction					
	(1)	(2)	(3)	(4)	(5)
Home	0.31***	0.28***	0.25***	0.19***	0.19***
	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)
HomeXSales	-0.21**	-0.19**	-0.19**	-0.19**	-0.19**
	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)
Away	0.06	0.08***	0.06**	-0.00	-0.00
	(0.04)	(0.03)	(0.02)	(0.01)	(0.01)
AwayXSales	-0.03	-0.02	-0.02	-0.02	-0.02
	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
Observations	162898	162898	162898	162898	162898
Number of Agencies	79	79	79	79	79
Agency FE	-	X	X	X	X
Years	-	-	X	X	X
Months	-	-	X	X	X
Week of Yr	-	-	-	-	X
Day of Week	-	-	-	X	X
Holidays	-	-	-	X	X

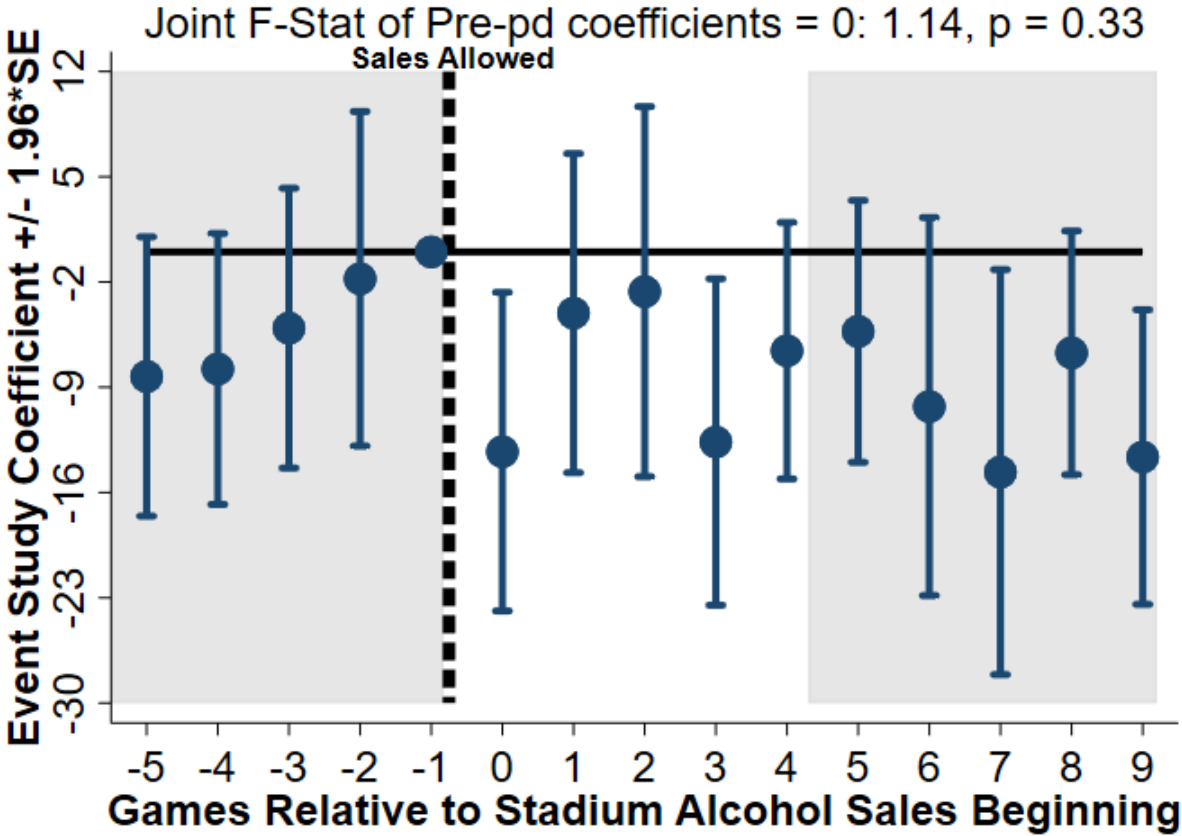
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors cluster-robust at the city-level in parentheses. All models estimated using Poisson conditional fixed effects models. Holidays are: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day.

Figure 1.3: Randomization Inference, Distribution of Z-Statistics From Home X Sales From Equation 1.1 Using Randomized Treatment Across Seasons and All Teams in Sample



Notes: Histogram displays the Z-statistic on the Home X Sales coefficient from Table 1.2, specifically the coefficient, β_3 from Equation 1.1, with the same fixed effects as column 4 of ???. No restrictions are imposed on the number of teams who receive treatment (i.e. Bernuolli sampling.) The treatment is randomized across time. The actual z-statistic is only greater than 13/1000 random samples, implying an exact Fisher p-value of 13/1000 = 0.013.

Figure 1.4: Time-Varying Treatment Effects of Sales on Home Game Crime Compared to Away Game Crime



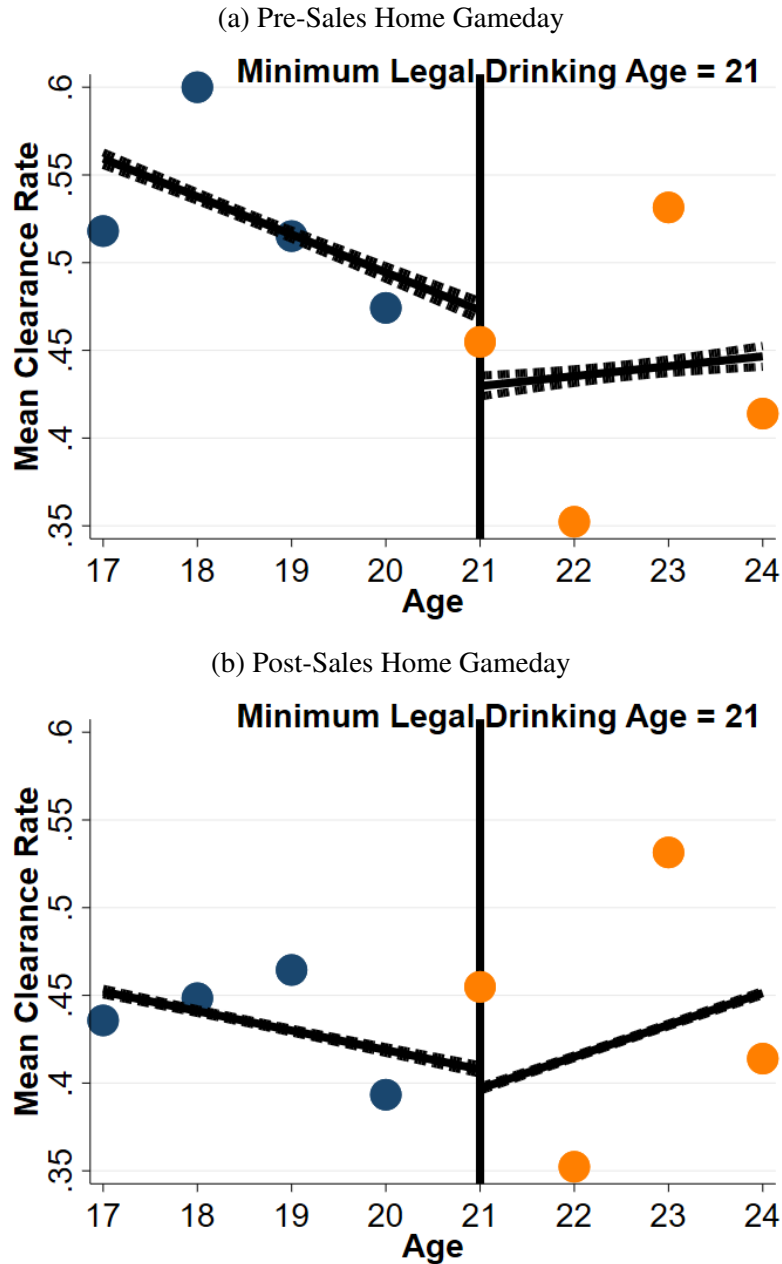
Notes: Coefficients are from the event study in Equation 1.3. Dependent variable is home crime - away crime. N = 551 event week-agency observations. Only teams that change included. Only gamedays included. Heteroskedasticity-robust standard errors used. Alternating between gray and white background shows the beginning of a new season.

Table 1.3: Investigating the Influence of Omitted Law Enforcement on Effect of Sales on Crime

	(1)	(2)
Home	0.07*** (0.01)	0.07*** (0.01)
HomeXSales	-0.09*** (0.02)	-0.08*** (0.03)
Away	0.00 (0.01)	-0.00 (0.01)
AwayXSales	-0.03 (0.02)	-0.02 (0.02)
Observations	131841	162898
OLS	X	X
Drop Missing Ratios	X	-

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors cluster-robust at the city-level in parentheses. All models include agency, year, month fixed effects. All models use indicator variables for day of week and holidays including: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day. All models estimated using OLS. Clearance rates are the dependent variable to measure law enforcement activity. Clearance rate = $\frac{\text{Number of Arrests}}{\text{Number of Reports}}$. Column 1 does not include observations where there are 0 arrests or 0 reports. Column 2 assumes clearance rate = 0 if there are no arrests.

Figure 1.5: How Enforcement Changes After Sales at the Legal Alcohol Access Age Threshold



Note: Panel A and B show the likelihood of being arrested, conditional on a report being filed by age. The red line vertical at age 21 represents the minimum legal drinking age cutoff. The likelihood of being arrested, clearance rate, has been used to proxy for enforcement intensity in prior literature (Heaton, 2012; Lindo et al., 2018).

Table 1.4: Effect of Alcohol Sales on Crime by Type Legal Alcohol Access Age

Panel A: Assaults									
	Aggravated Assault			Simple Assault			Intimidation		
	(1) ≥ 21	(2) < 21	(3) Both	(4) ≥ 21	(5) < 21	(6) Both	(7) ≥ 21	(8) < 21	(9) Both
Home*Sales	-0.28*** (0.08)	-0.11 (0.09)	-0.11 (0.10)	-0.32*** (0.09)	-0.22*** (0.08)	-0.22*** (0.08)	-0.06 (0.06)	-0.09 (0.07)	-0.09 (0.07)
Home*Sales*≥ 21			-0.17** (0.08)			-0.10** (0.05)			0.04 (0.08)
Observations	162898	162898	325796	162898	162898	325796	162898	162898	325796
Panel B: Bad Behaviors									
	Disorderly Conduct			Vandalism			Rape		
	(1) ≥ 21	(2) < 21	(3) Both	(4) ≥ 21	(5) < 21	(6) Both	(7) ≥ 21	(8) < 21	(9) Both
Home*Sales	-0.83*** (0.30)	-0.79** (0.39)	-0.79** (0.39)	-0.36*** (0.08)	-0.14 (0.16)	-0.14 (0.16)	-0.36** (0.18)	0.48* (0.26)	0.47* (0.26)
Home*Sales*≥ 21			-0.03 (0.20)			-0.22* (0.12)			-0.82*** (0.24)
Observations	162898	162898	325796	162898	162898	325796	162898	162898	325796
Panel C: Liquor Types									
	DUI			Liquor Law			Drunkenness		
	(1) ≥ 21	(2) < 21	(3) Both	(4) ≥ 21	(5) < 21	(6) Both	(7) ≥ 21	(8) < 21	(9) Both
Home*Sales	-0.20** (0.10)	-0.27** (0.12)	-0.29** (0.12)	-1.95*** (0.67)	-0.81*** (0.18)	-0.80*** (0.18)	-0.39 (0.26)	-0.47* (0.25)	-0.48* (0.25)
Home*Sales*≥ 21			0.09 (0.13)			-1.24** (0.60)			0.09 (0.28)
Observations	162898	162898	325796	162898	162898	325796	162898	162898	325796
Panel D: Falsification									
	Grand Theft Auto			Embezzlement			Forgery		
	(1) ≥ 21	(2) < 21	(3) Both	(4) ≥ 21	(5) < 21	(6) Both	(7) ≥ 21	(8) < 21	(9) Both
Home*Sales	0.08 (0.24)	-0.20 (0.14)	-0.21 (0.14)	-0.23 (0.18)	-0.68** (0.33)	-0.69** (0.33)	0.33* (0.19)	-0.36 (0.40)	-0.37 (0.40)
Home*Sales*≥ 21			0.29 (0.26)			0.46 (0.41)			0.72** (0.36)
Observations	162898	162898	325796	162898	162898	325796	162898	162898	325796

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors cluster-robust at the city-level in parentheses. All models estimated using Poisson conditional fixed effects models. Holidays are: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day. Columns 1, 4, and 7 subsample to crimes committed by those 21-23 years old. Columns 2, 5, and 8 subsample to crimes committed by those who are 18-20 years old. Columns 3, 6, and 9 use 18-23 year olds and estimate triple difference models.

Table 1.5: Effect of Alcohol Sales on Crime Victimization

Panel A: Total Victimization			
	(1) ≥ 21	(2) < 21	(3) Both
Home*Sales	-0.30*** (0.06)	-0.21*** (0.08)	-0.21*** (0.07)
Home*Sales*≥21			-0.10*** (0.04)
Panel B: Simple Assault			
	(1) ≥ 21	(2) < 21	(3) Both
Home*Sales	-0.37*** (0.11)	-0.23*** (0.07)	-0.23*** (0.07)
Home*Sales*≥21			-0.14** (0.06)
Panel C: Aggravated Assault			
	(1) ≥ 21	(2) < 21	(3) Both
Home*Sales	-0.32*** (0.09)	-0.16* (0.09)	-0.16* (0.09)
Home*Sales*≥21			-0.17** (0.08)
Panel D: Intimidation			
	(1) ≥ 21	(2) < 21	(3) Both
Home*Sales	-0.16* (0.09)	-0.21*** (0.05)	-0.20*** (0.05)
Home*Sales*≥21			0.04 (0.08)
Panel D: Rape			
	(1) ≥ 21	(2) < 21	(3) Both
Home*Sales	-0.29 (0.26)	-0.27 (0.27)	-0.27 (0.27)
Home*Sales*≥21			-0.03 (0.28)

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors cluster-robust at the city-level in parentheses. All models estimated using Poisson conditional fixed effects models. Holidays are: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day. Columns 1, 4, and 7 subsample to crimes committed by those 21-23 years old. Columns 2, 5, and 8 subsample to crimes committed by those who are 18-20 years old. Columns 3, 6, and 9 use 18-23 year olds and estimate triple difference models. All models include agency fixed effects, day of week indicators, year fixed effects, and month fixed effects. Column 3 includes full set of away, sales, and legal age interactions also.

Chapter 2

College Merit-Aid and Academic Renewal Thresholds: Evidence from Interactions of Discontinuous Rules

College graduates make \$1 million more in lifetime earnings on average than non-college graduates (Carnevale, Cheah, & Hanson, 2015); however, college costs remain a primary obstacle to college (Cowan, 2016; Solis, 2017). To combat financial constraints, college financial aid has grown substantially.¹ West Virginia spends \$40 million yearly (West Virginia Higher Education Policy Commission, 2014) on merit-based aid and Georgia had, from 1993-2011, 1,383,718 students receive merit aid.² Figure B.1 shows 27 states with merit-aid programs.³ Abbott, Gallipoli, Meghir, and Violante (2019) suggests that expanding ability-tested grants is superior to labor supply tax cuts or expanding student loans, but merit-aid is criticized for being ineffective (Ortagus,

¹College grant aid per full-time enrolled student has doubled from \$4,360 in 1998 to \$9,520 in 2018 (College Board, 2019).

²Georgia data accessed at: https://web.archive.org/web/20111006161755/https://www.gsfc.org/GSFCNEW/SandG_facts.CFM?guid=&returnurl=http://www.gacollege411.org/FinancialAidPlanning/HOPE_Program/Georgia_s_HOPE_Scholarship_Program_Overview.aspx. This citation required the use of waybackmachine.com.

³This is the classification given by Dynarski (2004), Heller and Marin (2004), The Brookings Institution, and State Administrative Websites. 9 states offer a generous large program. To be considered large, it means a large amount of students are covered and/or a large portion of expenses are covered.

Kelchen, Rosigner, & Voorhees, 2018) and a transfer to the wealthy (C. Cornwell & Mustard, 2007; Dynarski, 2004).

To inform debate about merit-aid and compare it to other interventions, it is crucial to get accurate treatment effects for the many outcomes merit-aid affects (C. Cornwell & Mustard, 2007; Cowan & White, 2015; Goodman, 2008; Scott-Clayton, 2011; Scott-Clayton & Zafar, 2019; Sjoquist & Winters, 2014, 2015b, 2015c). Since causal effects are used for comparing the welfare effects of government policies, accurate estimates have significant implications for understanding how merit-aid compares to other government policies (Hendren & Sprung-Keyser, 2020). A seemingly attractive research design to evaluate the causal effects of merit-aid is regression discontinuity (RD), because there are defined academic cutoffs which discontinuously change the probability of treatment. Using entry test (ACT, SAT) scores and/or high school GPA cutoffs is similar to test cutoffs used in the original RD study (Cook, 2008; Thistlethwaite & Campbell, 1960).

However, an important difference between current merit-aid research and Thistlethwaite and Campbell (1960) is that academic cutoffs for many merit-aid scholarships are known. Known academic cutoffs present a challenge to the RD design, because individuals can manipulate scores into receiving treatment and there is a large economic incentive to do so.⁴ Indeed, students strategically manipulate academic scores into being just high enough to receive treatment, by retaking entry exams such as the ACT, in Tennessee, Florida, and West Virginia (Bruce & Carruthers, 2014; Scott-Clayton & Zafar, 2019; Zhang, Hu, Sun, & Pu, 2016). This selection into treatment is not consistent with identifying continuity assumptions typically used for RD (Dong, 2018; J. Hahn, Todd, & Van der Klaauw, 2001).

This paper estimates the effects of West Virginia's merit-aid program, that at the time covered the entirety of students' tuition and fees, on reaching academic thresholds to renew the scholarship. Academic renewal thresholds have been an important design element of merit-aid programs (C. M. Cornwell, Lee, & Mustard, 2005; Jia, 2019) and are comparable to performance goal-setting schemes (Beattie, Laliberté, Michaud-Leclerc, & Oreopoulos, 2019; Clark, Gill, Prowse, & Rush,

⁴Two academic criteria are entry test scores and high school GPA. Entry tests can be retaken and high school GPA can be inflated through easier courses being taken or idiosyncratic high school decisions.

2020). West Virginia's merit-aid program, the Promise Scholarship, has known academic cutoffs; so, the RD design in this paper exploits variation in college scholarship receipt due to being born on opposite sides of the kindergarten entry cutoff date. This date is set by statewide compulsory schooling laws in place 14 years before the college scholarship begins in 2002. By being born on opposite sides of the cutoff, students enter in different kindergarten cohorts which affects their college cohort and eligibility to receive the scholarship.

This research design is less vulnerable to running variable manipulation, because birthdate is realized 16 years before the scholarship announcement and 19 years before it begins. Because the variables that lead to differences in scholarship receipt are realized 20 years before the scholarship begins, it is implausible that future students or their parents manipulated birthdates to increase their likelihood of becoming eligible for the scholarship. Since the cutoff determines both scholarship eligibility and age relative to the rest of ones' cohort, the prior year's date cutoff is used to examine and remove potential confounding due to relative age. Fusing compulsory entry laws as variation in education (Black, Devereux, & Salvanes, 2011; Clark & Royer, 2013) and a birthdate-based discontinuity in financial aid for college (Denning, 2019), this design uses early education laws as a natural experiment in college grant receipt.⁵ This design addresses two pervasive challenges, selection into treatment and sample, in the voluminous, active quasi-experimental literature on merit-aid program evaluation.

First, internal validity is investigated. The kindergarten cohort birthdate cutoff creates variation in college cohort and college scholarship receipt. Academically eligible students born slightly after the kindergarten cohort birthdate cutoff are 53.7 percentage points (pp) more likely to enter in the first eligible college cohort and 49.6 pp more likely to receive the scholarship than academically eligible students born slightly before the cutoff.⁶ Additionally, the treatment and control groups are reasonably similar, because density tests suggest birthdate is unmanipulated (McCrary, 2008) relative to the the prior year's birthdate cutoff (Grembi et al., 2016).

⁵A direction for future research in financial aid noted by Dynarski and Scott-Clayton (2013) is analyzing how college financial aid interacts with other programs.

⁶Students who are sophomores in college when the program begins (i.e. begin college in 2001) are not retroactively eligible.

This alternative research design reveals new insights into the effects of merit-aid scholarships on academic outcomes. There is no difference in the likelihood of reaching pace-based (credit hours) or distinction-based (GPA) performance goals. Point estimates indicate students who receive the scholarship are 24 to 30 pp more likely to hit both the credits and GPA renewal thresholds in their first year of college.

There are heterogeneous effects on exceeding academic renewal thresholds by gender and academic credentials. Results suggest that those above the median ACT score are more likely to reach the credits threshold than those below it. Also, students below the median high school GPA are more likely to reach the GPA thresholds. Finally, there is no difference between men and women for exceeding renewal credits in each of the first 3 years; however, men are more likely to complete the GPA requirement in each of their first 3 years, but women are not.

One way these estimates may diverge from a causal treatment effect of the scholarship is due to relative age confounding at the cutoff ([D. Deming & Dynarski, 2008](#)); however, relative age effects are found to be empirically negligible in this context.⁷ Relative age changes discontinuously at the pre-scholarship cutoff and scholarship eligibility does not. Specifically, students are 14 pp less likely to be male, score 0.84 points lower on the ACT Science subsection, and score 83.31 points higher on their combined SAT scores at the pre-scholarship cutoff, which is hardly indicative of systematic relative age effects on pre-treatment covariates.

Second, exceeding the birthdate cutoff only has positive effects on 1 of 4 group's academic performance in college, those who are academically eligible and born just beyond the post-scholarship cutoff (i.e. those who become eligible for the scholarship by exceeding the cutoff.)⁸ Finally, subtracting the intent-to-treat effects of relative age from being just beyond the pre-scholarship cutoff from the local average RD treatment effect, using a difference in discontinuities estimate ([Grembi et al., 2016](#)), has nearly no effect on the point estimates. These estimates are also not vulnerable to concerns about how mothers' characteristics differ across seasons of birth, because there is no

⁷For a review of relative ages effect on mostly education outcomes, see [Peña \(2017\)](#).

⁸There is no effect for those who are academically ineligible at either cutoff and there is no effect for academically eligible students born just beyond the pre-scholarship cutoff.

reason to believe there is a difference in mothers' characteristics on either side of the pre-scholarship cutoff compared to the post-scholarship cutoff (Buckles & Hungerman, 2013).

Students must finish both requirements to renew the scholarship and research in Georgia has found merit-aid reduces academic performance (C. M. Cornwell et al., 2005), so the large discrepancy in prior work (Scott-Clayton, 2011) warrants additional attention. The most closely related prior work finds the merit-aid scholarship increases the likelihood of hitting pace-based (credit hours) performance was 24 pp and distinction-based (GPA) was 8 pp (Scott-Clayton, 2011). The results in this paper suggest that this difference is a limitation of methods used, rather than a substantive behavior difference.⁹ Policies based on this apparent difference in behavioral responses to different measures of academic performance are likely misguided. Furthermore, it supports the conclusions of C. M. Cornwell et al. (2005), which suggests that the reason Georgia's merit-aid reduced academic performance was it did not limit the number of semesters that aid could be used.

These results are important given the growth of interest in behavioral interventions within education (Gneezy, Meier, & Rey-Biel, 2011; Lavecchia, Liu, & Oreopoulos, 2016; Levitt, List, Neckermann, & Sadoff, 2016). By setting a threshold for renewal, the program provides a reference point (Kahneman & Tversky, 1979) and this point conceivably represents a goal. Those who fail out of college are less likely to set goals (Beattie et al., 2019).

However, the performance goal set by the minimum renewal threshold might be expected to be ineffective without the incentive of financial aid. Without financial incentives, performance goals have been found to be less effective than task-based goals for college students (Clark et al., 2020). Similarly, without accompanying support, financial aid for performance has also been found to be less effective (J. Angrist, Lang, & Oreopoulos, 2009). These results suggest that a large financial incentive for completing a minimum level of academic performance can effectively increase academic performance. Indeed, if students complete the 30 credits renewal requirement each year, they will have obtained the minimum number of credits, 120, required to graduate in 4 years. While there are justified targeting concerns with merit-aid, it is effective at bolstering

⁹Only the widest bandwidths with uniform kernels are able to replicate the differences across GPA and credits thresholds.

academic performance in West Virginia.

Beyond policy relevance of merit-aid, this research fits the broad direction of applied microeconomics. The alternative source of variation is consistent with estimating treatment effects using natural experiments (J. D. Angrist & Pischke, 2010).¹⁰ In contrast to methodological approaches to addressing heterogeneity in treatment effects (Hsu & Shen, 2019), monotonicity (Bertanha & Imbens, 2019), or extrapolating away from cutoffs in RD designs (J. D. Angrist & Rokkanen, 2015; Dong & Lewbel, 2015), this paper estimates a “more average” effect using a different natural experiment. By using a birthdate running variable, the estimated effects are not local to an academic cutoff; estimates are local near a date, making generalizations away from the cutoff more likely to be appropriate.

Exact birthdate is continuous (while ACT is more discrete) which is consistent with identifying assumptions typically used (J. Hahn et al., 2001). The continuous running variable is less vulnerable to error due to specification bias caused by discrete variables not being arbitrarily close to the cutoff (Lee & Card, 2008) or choice of standard errors (Kolesár & Rothe, 2018). The estimates may be more heterogeneous, because potential students have less exact control of birthdate compared to entry test (ACT/SAT) scores. Higher running variable variance (as is arguably the case with a birthdate running variable compared to an entry test score) implies a greater mixture of students affected by the birthdate cutoff (Bloom, 2012; Lee, 2008) meaning more groups represented by estimates.

Finally, the empirical strategy in this paper, using Kindergarten birthdate cutoffs to investigate a program beginning with a later grade, has started to become more regularly applied (Canaan, 2020; Takaku & Yokoyama, 2021). This strategy is appealing, because it approximates a natural experiment. It compares students who are born a few days apart and that unknowingly at the time sorted them into eligible and not eligible groups. Also, it can be widely applied, because compulsory laws are widespread and many programs begin with a specific cohort. Developing this strategy can result in a new, rigorous method for evaluating causal effects of programs in which

¹⁰The new source of variation provides a new way to form a counterfactual to solve the fundamental problem of causal inference (Holland, 1986).

individuals are treated beginning with a specific cohort and compulsory schooling entry birthdate cutoff laws existed when said individuals entered school.¹¹

This paper extends prior work with this empirical strategy in three ways. First, it uses exact date of birth as the running variable while previous work has been limited to birth month-year. Second, it explicitly models the process for how kindergarten birthdate cutoffs can be used to estimate effects of programs that begin in a grade beyond kindergarten, noting specifically that this strategy depends on grade progression. This basic model can be used to inform assessments of how appropriate this strategy is for other contexts and/or programs and who the compliers are (J. D. Angrist, Imbens, & Rubin, 1996). Finally, this is the first paper to apply this empirical strategy to the debate surrounding the effects of merit-based aid for college.

The following section discusses related prior research. Then, the programs' rules and resulting variation are formally modeled. Next, the variation is validated as being viable and internally valid. Finally, methods and results are presented, followed by a brief conclusion.

2.1 Related Literature

This section begins by reviewing students' response to incentives created by the academic renewal thresholds that are common for merit-aid programs. Then, the research designs that are commonly used to estimate effects of merit-aid are outlined. The section concludes with comments on alternative strategies to address the econometric issues caused by students manipulating academic scores into treatment.

2.1.1 Behavior Responses to Academic Renewal Thresholds

There are several important aspects of merit-aid program design, an important one is academic renewal thresholds. There are discrepancies across different states' merit-aid programs in how

¹¹Also note that in general applied economics has recently seen a massive increase in exploiting quasi-random variation from historical sources, such as exploration routes (Duranton & Turner, 2011), colonization (Acemoglu, Johnson, & Robinson, 2001), bombing (Dell & Querubin, 2018), and policy changes (Lleras-Muney & Shertzer, 2015).

merit-aid affects pace-based academic performance. In Georgia, students were allowed to use their aid for 180 credits and that reduced the likelihood of completing a “full-load” by 9.3% and reduced the credits completed by 1 (C. M. Cornwell et al., 2005). When there is a credits time limit and academic renewal thresholds required for renewal, students would drop courses to ensure the academic thresholds were exceeded.

Contrast Georgia’s credit limit with West Virginia, where students can only use aid for 8 contiguous semesters. When the limit on funds is determined by semesters, not credit hours, students are much more likely to reach academic renewal thresholds. The puzzle that remains for academic renewal thresholds in West Virginia is why are students 8 pp more likely to complete GPA requirements and 24 pp more likely to complete credits requirements (Scott-Clayton, 2011)?¹² After all, students need to hit both renewal thresholds to renew the scholarship. Furthermore, does changing the time limit for receiving merit-aid from credits to semesters switch the sign from negative to modestly positive (8 pp) or to a large positive (24 pp) or is it actually different for different academic thresholds?

Another behavioral response to academic renewal thresholds is major choice. Students might choose easier majors to reduce the risk of meeting academic renewal thresholds. Students choose easier majors in Georgia (Sjoquist & Winters, 2015a) and nationwide (Sjoquist & Winters, 2015b), to make meeting the thresholds easier. However, students have not been shown to be less likely to choose STEM majors in West Virginia (Scott-Clayton, 2011).

2.1.2 Research Designs

There are two primary sources of variation that have been used to estimate the impacts of merit-aid: cohort-based and academic eligibility based. The academic eligibility design exploits small differences in test scores and/or GPA along with academic eligibility cutoffs. However, in many states these academic eligibility cutoffs are known, leading to students manipulating academic scores into

¹²In West Virginia, students are found to increase their total credits in their first year by 1.8 and there are no significant impacts on GPA (Scott-Clayton, 2011).

being eligible which is inconsistent with the oft-used continuity assumptions for identification.

The cohort-based variation exploits the fact that students who enter college before merit-aid programs begin are not eligible retroactively. This leads to two different designs, which depend on data availability. The first design estimates differences-in-differences models, where the control group is states that do not begin the program and estimated effects are intent-to-treat (ITT) if individual-level data is used.

The ITT effects can be scaled to a local average treatment effect if it is observed who actually receives the scholarship as in [Scott-Clayton \(2011\)](#) and [Scott-Clayton and Zafar \(2019\)](#). In this setup, entering in an eligible cohort is used as an instrumental variable for scholarship receipt. Scholarship receipt is then used to explain individual-level outcomes. One limitation of this approach is that the “enter in an eligible cohort” instrumental variable is not necessarily defined at the individual level, it is defined at the cohort level. Seen from this perspective, a cohort-level instrumental variable is being used to explain individual-level outcomes which may be less than ideal.¹³ Furthermore, the discrete instrument makes it unclear who the compliers are.

2.1.3 Previous Strategies for Addressing Academic Score Manipulation

In order to estimate causal effects of merit-aid by addressing academic score manipulation, there are two strategies that have been seen in the literature so far. The first strategy is to use the first entry test score instead of the best entry test score ([Bruce & Carruthers, 2014](#)). There is no manipulation in the first score, because students have not retaken the test yet. The limitation of this approach is that it requires additional data on whether the score recorded is the first or best, which may not be possible with many datasets.¹⁴

The second strategy for addressing academic entry manipulation is to use unknown academic

¹³It is shown in [Scott-Clayton \(2011\)](#) that collapsing to cohort-level means, the effect remains statistically significant.

¹⁴In fact, it is not explicit in studies that use administrative state datasets whether the test scores are first score or best score or something else like combined across multiple attempts. It could be that some schools report the best score and others report the first score. Combining across attempts-a process known as superscoring- is how this university assigns institutional scholarships.

cutoffs. If the academic scores are not known, students cannot sort into treatment by manipulating their scores just beyond the threshold. One state with ex ante unknown cutoffs for merit-aid is Massachusetts (Cohodes & Goodman, 2014; Goodman, 2008). The limitation of using ex ante unknown cutoffs is that it limits the states that can be studied. Limited states with unknown academic cutoffs forces researchers to argue about whether results in Massachusetts are externally valid for states like West Virginia. This may be a difficult position, because West Virginia's population likely differs substantially in relevant ways from Massachusetts. Even still, regression discontinuity estimates from academic qualification thresholds may not generalize to students far away from cutoffs without additional strong assumptions (J. D. Angrist & Rokkanen, 2015).

2.2 Data

The data is student-level of the 2000-2002 cohorts at a large, public university in West Virginia.¹⁵ The sample consists of only in-state students who are first-time freshman. Exact birthdate, gender, race, high school GPA, entry test scores, and academic outcomes up to the third year of college are observed.¹⁶

Scholarship Eligibility The data has information on 3 requirements required to be eligible for the scholarship:

1. Earn at least a 3.0 GPA in high school,
2. ACT Composite score ≥ 21 OR SAT Combined Score $\geq 1,000$,
3. One must be a college freshman in the 2002 cohort or after.

As the program is merit-based, 1 and 2 are requirements related to academic proficiency. To

¹⁵After multiple rounds of data requests over multiple years, the West Virginia Higher Education Policy Commission (WVHEPC) ultimately denied the data request for statewide data.

¹⁶Table B.1 presents summary statistics of the sample.

simplify, requirements 1 and 2 are combined into a single variable, academic eligibility, defined as

$$Acad. Eligible_i = 1[(HS GPA \geq 3) \cap (ACT Comp \geq 21 \cup SAT Comb \geq 1,000)],$$

in which *HS GPA* stands for high school GPA, *ACT Comp* stands for ACT composite score, and *SAT Comb* stands for SAT combined score. The academic requirements change for subsequent cohorts, but the change is not relevant for this analysis, since there is only 1 eligible cohort used.¹⁷

Exact Birthdate and Cutoffs Exact birthdate is observed. According to standard practice, it is centered so that day = 0 is the first birthdate that is expected to be discontinuously more likely to receive the scholarship. The cutoff at day = 0 is called the post-scholarship cutoff, because being born on or after this date makes one more likely to receive the scholarship, as detailed in Section 2.3. Furthermore, students born at day = -365 are more likely to be the oldest in their cohort, but not more likely to receive the scholarship. Because it is a cutoff that does not determine scholarship receipt, it is referred to as the pre-scholarship cutoff throughout this paper.

2.2.1 Outcomes: Academic Performance Based Renewal

The dependent variables, y , are binary indicators for whether a student finished the minimum required credits or GPA in their first year. Students must finish at least 30 credits in each year to renew their scholarship.¹⁸ In their first year, they must obtain at least a 2.75 GPA.¹⁹ By focusing on the first year, it ensures that results are not driven by selection out of the sample as students drop out of college. While continuous outcomes are interesting, binary outcomes for exceeding renewal requirements have been used before (Scott-Clayton, 2011), performance based funding is criticized (Ortagus et al., 2018), performance based goals have been shown to be less effective than task based goals (Clark et al., 2020), and the binary indicators have implications for program

¹⁷There are also subscore requirements, minimum scores on each section- math, science, reading, verbal, etc.- that are not incorporated into the analysis.

¹⁸They have the Fall, Spring, and Summer semester to do so.

¹⁹For every year afterwards they must obtain at least a 3.0 GPA.

design (C. M. Cornwell et al., 2005).

Outcomes of Interest By Cohort

Figure B.3 visualizes how a differences in differences analysis would estimate effects of the program by plotting outcomes by cohort and academic eligibility.²⁰ There is a positive effect for both likelihood of completing credits (Figure B.3a) and GPA (Figure B.3b). However, a worrisome problem with estimating treatment effects this way is that, due to manipulation of test scores, those who would have had a 20, now have a 21 and find themselves in the academically eligible group. Furthermore, estimating the treatment effects this way is unable to leverage the data fully, specifically it does not use the information of who received the scholarship which is in the data.

Possible Biases From Academic Score Manipulation The effect of this academic score sorting is ambiguous on the direction it biases the treatment effects. On one hand, students may study more and thus increase their scores which may bias estimates upward. However, students did not have this incentive before and so students would not have retaken tests. Thus they would have still had a 20 in cohorts before the program begins. If they sort into treatment, but are not more academically able, then the treatment effects of the scholarship are biased downwards, because the academically eligible are less able on average. The next section describes an alternative source of variation to circumvent relying on mismeasured and manipulated academic eligibility as variation.

2.3 Kindergarten Cohort, College Cohort, and the Promise Scholarship

This section discusses the discontinuous rules and related factors that lead to discontinuous college scholarship receipt at the kindergarten birthdate cutoff, 14 years before the scholarship begins. The main insight builds on scholarship receipt depending on college entry cohort. Specifically, under

²⁰It would be possible to use out of state students as a comparison group like Scott-Clayton (2011), but I do not have data on out of state students.

a mild grade progression assumption, college entry cohorts depend predictably on kindergarten entry cohorts.

If college entry cohort depends predictably on kindergarten cohort, and the scholarship depends on college entry cohort, then this is an alternative source of variation to estimate effects of merit-aid programs on individuals. Conveniently, kindergarten entry cohort depends discontinuously on birthdate in relation to kindergarten birthdate cutoff due to state compulsory entry laws. While at least two other papers ([Canaan, 2020](#); [Takaku & Yokoyama, 2021](#)) have used this logic to investigate treatments that begin with a specific cohort, this is the first paper that makes the underlying theory responsible for the discontinuous variation mathematically explicit.

Equations [2.1](#), [2.2](#), and [2.3](#) outline the system of equations which leads to the variation in scholarship receipt exploited in this paper,

$$\textit{Promise Scholarship} = f(\textit{College Cohort}), \quad (2.1)$$

$$\textit{College Cohort} = g(\textit{Kindergarten Cohort}), \quad (2.2)$$

$$\textit{Kindergarten Cohort} = h(\textit{Birthdate Relative to Cutoff}). \quad (2.3)$$

Equations [2.1](#) and [2.3](#) are known, discontinuous rules about scholarship eligibility and kindergarten cohort entry respectively. [Equation 2.2](#) is related to grade progression. By substitution,

$$\textit{Promise Scholarship} = f(g(h(\textit{Birthdate Relative to Kindergarten Cutoff}))). \quad (2.4)$$

Each of these equations is discussed below.

First, [Equation 2.1](#) represents the rule that to receive the scholarship, students must enter college after the scholarship program begins. [Equation 1B](#) expands [Equation 2.1](#) into its discontinuous

form,

$$P(\text{Scholarship} = 1) = \begin{cases} q' + k' & \text{if Entry Cohort} \geq 2002 \\ q' & \text{if Entry Cohort} < 2002. \end{cases} \quad (1B)$$

The constant k' can be considered how important being in an eligible entry cohort is for scholarship receipt. Prior research suggests $k' > 0$, meaning that entry cohort is an important eligibility factor determining scholarship receipt (Scott-Clayton, 2011; Scott-Clayton & Zafar, 2019).

Next, Equation 2.3 represents the compulsory kindergarten entry birthdate cutoff rule. This is a rule set by laws that were in place in 1988 in West Virginia. Students who have not turned 5 by this date, September 1, are required to enter kindergarten in the next cohort, effectively requiring a birthday, BD , of a certain date. Equation 3B expands Equation 2.3 into its discontinuous form for the 1989 kindergarten cohort,

$$P(\text{Kindergarten Cohort} = 1989) = \begin{cases} p' + c' & \text{if } BD \geq \overline{BD}_{y=1988} \\ p' & \text{if } BD < \overline{BD}_{y=1988}. \end{cases} \quad (3B)$$

The equation suggests that with a large enough c' , students are discontinuously more likely to enter kindergarten in 1989 if born after the cutoff date, $\overline{BD}_{y=1988}$, for entering kindergarten in 1988. The only way c' could equal 0, invalidating the design, is that the cutoff laws are not enforced. In practice, the cutoff is imperfectly enforced and costly to circumvent, as discussed in Section B.2.²¹

Finally, Equation 2.2 represents a grade progression function. In order to link Equation 2.1 to Equation 2.3 in a systematic and predictable way for estimating effects of the college scholarship, Equation 2.2 must take a specific functional form for a large enough share of the population. The assumed functional form of g is,

$$\text{College Cohort} = \text{Kindergarten Cohort} + 13,²² \quad (2B)$$

²¹Since kindergarten entry cohort is not observed, it is impossible to estimate c' (the degree of imperfect enforcement) using the available data.

²²There are 13 grades between kindergarten and college. For a program beginning with a different year, 13 would

that is students finish one grade per year (kindergarten through 12 grade) and then enter college. This is a reasonable assumption for the purposes of estimating the effects of merit-based aid, because it is unlikely that students who receive merit-based aid are being held back.

Kindergarten entry cohort and grade progression are not observed, but the assumption in [Equation 2B](#) can be supported by the data. Under the one grade per year progression assumption in [Equation 2B](#), [Equation 3B](#) implies,

$$P(\text{College Cohort} = 2002) = \begin{cases} p + c & \text{if } BD \geq \overline{BD}_{y=1988} \\ p & \text{if } BD < \overline{BD}_{y=1988}. \end{cases} \quad (2.5)$$

If $c \neq 0$, then it strongly suggests that [Equation 2B](#) is realistic for a large enough share of the population. The reason c' , which previously represented only how strictly the compulsory law was enforced, is now c , is because c also depends on how large of a share of the population follows the one grade per year assumption. A final practical matter which might push c towards 0, to the detriment of the design, is immigration to West Virginia from other states, with different cutoffs, in grade school.^{23,24,25} Since beginning in the eligible year matters ([Scott-Clayton, 2011](#); [Scott-Clayton & Zafar, 2019](#)), [Equation 2.5](#) implies

$$P(\text{Scholarship} = 1) = \begin{cases} q + k & \text{if } BD \geq \overline{BD}_{y=1988} \\ q & \text{if } BD < \overline{BD}_{y=1988}. \end{cases} \quad (2.6)$$

If $c \neq 0$ and beginning in the first year of the program is an important enough requirement, then it

be a different number. For example, if the program began in 6 grade like [Canaan \(2020\)](#), then 13 would be 6.

²³This could be remedied if state of birth was observed; however, migration will cause estimates to be conservative. It is likely this is not much of a problem, given the average annual growth rate of the population of West Virginia from 1950-2000 was -0.12%.

²⁴The scholarship has a rule that a student must gain resident status at least 1 year in advance, so students cannot migrate into West Virginia to receive the scholarship.

²⁵([Lleras-Muney, 2002](#), pg. 407) notes a similar point in footnote 36, going further to point out that inter-state mobility was probably not related to compulsory entry laws, which is likely also the case for this paper. Further, [Lleras-Muney \(2002\)](#) notes that mobility is low in the US in the period of study ([Card & Krueger, 1992](#)). Mobility into West Virginia is likely also low. Finally, it is highly unlikely that someone moved into West Virginia early in grade school to receive the scholarship.

is expected that $k \neq 0$. Without additional restrictions, $c \neq k$, because academic eligibility is also an important criterion for eligibility.

Finally, it is worth noting that $p = 0$ and $c = 1$ and/or $q = 0$ and $k = 1$ are not necessary (or sufficient) conditions for this design to be valid. Since it is likely that $(p + c) - p < 1$, a fuzzy regression discontinuity design can be used instead of a sharp design. All the terms in Equations 2.5 and 2.6 are observed, so p , c , q , and k are non-parametrically estimated, around $\overline{BD}_{y=1988}$ (and $\overline{BD}_{y=1987}$), in Section 2.4.1.

2.3.1 Scholastic and Birthdate Cutoff Comparisons

For a 10% random sample, Figure 2.2 plots how scholarship receipt depends on birthdate and academic credentials. There are few orange o's and many blue x's to the left of the birthdate cutoff and vice-versa to the right of the cutoff in Figures 2.2a and 2.2b. This suggests that c and $k > 0$.

Figure 2.2a compares the ACT cutoff to the birthdate cutoff. There are less individuals just below the ACT cutoff for those born after the birth cutoff, which is consistent with findings in Scott-Clayton (2011) and Scott-Clayton and Zafar (2019). There are a limited number of observations below the cutoff and many above the cutoff which is not ideal for RD, because it suggests students manipulate scores into treatment.²⁶ The limited number of observations near the ACT cutoff is also not good for RD, because it needs many observations near the cutoff.

Figure 2.2b compares the HS GPA cutoff to the birthdate cutoff. There are very few observations near the GPA cutoff before or after the birthdate cutoff, meaning it is unlikely the GPA cutoffs are appropriate for use in an RD design. There are many students who have near perfect/perfect high school GPAs on either side of the date cutoff. This shows the advantage of using a non-academic cutoff, it is possible to compare students of similar academic ability before and after the cutoff.²⁷

²⁶A formal investigation of density on opposite sides of the cutoff is performed in Section 2.4.2.

²⁷One more interesting takeaway message from Figure 2.2b is there is only 3 students who have a HS GPA below 2.5 born after the cutoff date. There are 6 students below 2.5 in (-730, -365) and 9 students below 2.5 in (-365,0). In the full sample, these numbers are 93/1801 students and in (-365,0) there are 93/1796, in (0, 365) there are 59/1868. This is likely related to high school grade inflation and/or crowd out of the less academically able students (Bound &

Threats to Identification There are two main threats to identification. The first is relative age effects, which is addressed by using the prior year’s cutoff which does not affect scholarship receipt. Relative age is estimated using the pre-scholarship cutoff and found to be empirically negligible. While students and parents can choose their birthdate in relation to the compulsory kindergarten cutoff, it is impossible they choose their birthday to be eligible for the scholarship. The second threat is there being another policy that begins with the cohort that begins Kindergarten in 1989. A search for such policies finds no policies that affect those discontinuously on either side of the birth cutoff.²⁸

2.4 Internal Validity of Research Design

This section investigates internal validity of using the proposed kindergarten cutoff as variation in scholarship receipt. First, the first stage is tested, specifically examining whether small differences in birthdate lead to different cohort year of entering college (c' in Equation 2.5) and differential likelihood of scholarship receipt (k' in Equation 2.6.) Second, implications of the identification assumption are examined in two ways. Density tests investigate smoothness in of number of observations at the date cutoff. Second, in the vein of a randomized controlled trial (RCT), pre-treatment covariate balance at the birthdate cutoffs are examined.

2.4.1 First Stage

The logic underlying the research design is that birthdate relative to kindergarten entry birthdate cutoffs predicts college entry cohort. Since scholarship eligibility is cohort-based, the birthdate cutoff also leads to variation in scholarship receipt. This section, describing Figure 2.3, tests whether birthday relative to the kindergarten birthdate cutoff causes differences in college entry cohort and scholarship receipt.

Turner, 2007).

²⁸These are discussed in detail in Section B.3.

Effect of Kindergarten Birthdate Cutoff on College Entry Cohort

The proposed research design requires a discontinuous increase in likelihood of beginning college 13 years later at the kindergarten entry birthdate cutoffs for beginning kindergarten. [Figure 2.3a](#) plots the likelihood of entering college in 2001, which is 1 cohort prior to the scholarship beginning. The cutoff is at birthdate = -365, because this pre-scholarship cutoff is on the same exact date in the prior year.

Students born just to the left of the cutoff in [Figure 2.3a](#), who are therefore expected to enter college in 2000 if they finish one grade per year, have an approximately 20% chance of entering college in 2001. As the students birthdates are further to the left of the cutoff, it becomes less likely that they enter in 2001, eventually reaching 0 after a few months. The leading explanation is that students just at the cutoff are more likely to successfully evade the cutoff in some way, so they can enter in a later kindergarten cohort.

Students born just to the right of the cutoff in [Figure 2.3a](#), who are therefore expected to enter college in 2001 if they finish one grade per year, have an approximately 70% chance of entering college in 2001. This probability grows as students are born further to right, until reaching a 100% chance after a few months. This probability falls again after a few months, because the data also contains students who enter college in 2002. Using a flexible local-linear regression with a triangular kernel and MSE-optimal bandwidths to estimate the difference in the conditional expectation function on either side of the cutoff, there is a difference of 52.6 percentage points at the cutoff.

[Figure 2.3c](#) inspects whether the kindergarten cutoff at birthdate = 0 determines whether a student enters college in 2002, the first year of the scholarship. This is the expected college cohort given expected kindergarten cohort and regular grade progression. Just like at the prior cutoff, about 20% of students to the left of the cutoff enter later than expected and about 75% enter in their expected cohort.²⁹ Using the same local-linear method for calculating the difference, the

²⁹Because there is no data on the 2003 cohort, the probability of entering in 2002 stays at 1 and does not exhibit the same reduction in probability far from the cutoff like in [Figure 2.3a](#).

kindergarten entry cutoff leads to an approximately 53.3 pp discontinuity in college entry cohort for entering in the first eligible college cohort. That is, $c = .533$.

No Evidence Students Delay College Entry One concern is that, because the scholarship is announced in 1999 and the program begins in 2002, students have an incentive to delay the beginning of college; however, this is unlikely for two reasons. First, the optimal decision on delay depends on comparing the value of the scholarship to the cost of delaying entering the labor market after college by 1 year as discussed in [C. M. Cornwell et al. \(2005\)](#). Second, the year before the scholarship begins has a 52.6 pp difference at the cutoff, while the year after the scholarship has a 53.3 pp cutoff which are not statistically different. Also, academically eligible students, who the strongest incentive to delay and are shown in [Figure 2.3e](#), are the same as the aggregate sample, with a 53.7 pp difference in entering in the expected cohort at the cutoff.

Effect of Kindergarten Birthdate Cutoff on Scholarship Receipt

[Figure 2.3b](#) shows the effect of exceeding the kindergarten birthdate cutoff, which by grade progression would have students enter in 2001 which is 1 year before the scholarship begins, on scholarship receipt. Students born immediately to the left and right of this cutoff do not receive this scholarship. This is expected, because being born after the cutoff leads one to be expected to enter college in the cohort before the scholarship begins.

[Figure 2.3d](#) shows the effect of being born to the right of the kindergarten birthdate cutoff, which by grade progression would have students enter in 2002 which is the first year the scholarship begins, on scholarship receipt. Students to immediately to the left of the cutoff receive the scholarship with about a 10% probability, while students to the right of the birthdate cutoff receive the scholarship with about a 50% probability.³⁰ Similar to the probability of entering in a given cohort, the probability falls to the left of the cutoff as get farther from it and increases to the right of the cutoff as one gets farther from it. Using the same local-linear methods, the difference in

³⁰There is substantial variance in the probability of receiving the scholarship to the right of the cutoff, with 40% in the lowest bin receiving and over 80% in the highest bin having receiving it.

scholarship receipt at the cutoff is 35.5 pp, that is $k = .355$, which is adequate in terms of a first stage effect on likelihood of receiving treatment. Also $c > k$, because [Figure 2.3d](#) does not account for a the academic eligibility criterion for merit-based aid.

[Figure 2.3f](#) repeats [Figure 2.3d](#) for students who are academically eligible. Students immediately to the left of the cutoff receive the scholarship with 20% probability, while those to the right receive it with 70% probability.^{31,32} The difference in scholarship receipt, k , at the cutoff is 49.6 pp, which is much closer to the 53.7 pp difference in [Figure 2.3e](#).³³ After accounting for academic eligibility, $c \rightarrow k$.³⁴ In summary, the kindergarten cutoff is a strong enough predictor of college cohort ($c > 0$) and scholarship receipt ($k > 0$) to be used in a RD design.

2.4.2 Density Around ACT Composite Score and Birthdate Cutoffs

The identifying assumption for the fuzzy RD research design used in this paper is local smoothness of potential outcomes at the cutoff ([J. Hahn et al., 2001](#)).³⁵ This assumption cannot be directly tested; however, one way to support the viability of this assumption is to test the density of the number of observations, around the birthdate cutoff ([McCrary, 2008](#)). The null hypothesis of the density test is a smooth distribution around the cutoff, so failing to reject is implicit support of the identifying assumption ([McCrary, 2008](#)).

The intuition is that if the distribution of observations is not smooth around the cutoff, then students who manipulated the running variable may be different from those who didn't (observably or unobservably). If students on opposite sides of the cutoff differ in their ability to manipulate the running variable, then it may inappropriate to assume that one side is an appropriate counterfactual

³¹Also, the variance on the right side of the cutoff decreases by including information on academic eligibility. Whereas there was a range of about .4, there is now only a range of about .25.

³²The probability of receiving the scholarship does get close to 1 after enough dates after the cutoff. It may not get all the way to 1, due to other types of aid replacing the Promise scholarship.

³³Figures [B.4](#) and [B.5](#) show that those to the right of the cutoff receive at least 1 additional year of the scholarship at the cutoff, because some students lose their scholarship ([Carruthers & Özek, 2016](#)).

³⁴The reason $c \neq k$ is due to other unobserved tertiary factors, such as ACT subscore and HS GPA subject requirements.

³⁵Other identifying assumptions include local randomization ([Cattaneo, Frandsen, & Titiunik, 2015](#)), which allows the use of randomization inference methods for regression discontinuity.

for the other. A failure to reject is evidence students are not manipulating the running variable into being eligible for treatment.³⁶ Ex-ante, this is a believable proposition, because students and parents did not know that being born slightly later would lead to being eligible for the scholarship. Something that could cause this test to fail is an increase in enrollment, which has been found in prior research (C. Cornwell, Mustard, & Sridhar, 2006; Scott-Clayton, 2011). However, to fail the test would require that enrollment increase right at the cutoff, relative to the pre-scholarship cutoff (Grembi et al., 2016).

ACT Composite Score Figure 2.4a shows there is a relatively smooth distribution of observations around the ACT cutoff value of 21 in the pre-scholarship cohorts. There are nearly the same number of students (about 280) who score 20 as 21 and less students score 23 and 24 than 19 or 20. After a score of 21, there is a steady decline in the number of people who receive each subsequent score. Figure 2.4b shows there is no longer a smooth distribution of ACT scores around the cutoff value of 21 for the first cohort of students who can receive a Promise scholarship if they get at least a 21.³⁷ In the first cohort of the scholarship beginning more students receive a 21-24 than a 19 or 20. This is unsurprising, because students know the cutoff and there is a large economic incentive to retake tests until they exceed it. The findings in Figures 2.4a and 2.4b replicate the ACT score manipulation in Scott-Clayton (2011) in a subsample of the state-wide data. The result suggests that using ACT score as a cutoff in an RD design is not consistent with the identification assumption that potential outcomes are smooth locally (J. Hahn et al., 2001).

Birthdate

Figures 2.4c and 2.4d show the birthdate density around the pre-scholarship and post-scholarship kindergarten cutoffs. There are 26 bins that contain 14 days (2 weeks) on either side of the cutoffs. In contrast to Figures 2.4a and 2.4b, there are many more bars in total, because the underlying

³⁶On the other hand, rejecting the null hypothesis is not a sufficient condition for running variable manipulation.

³⁷A formal density test at the ACT cutoff is not presented here, because the discrete running variable, ACT composite score, leads to econometric issues with the test (Frandsen, 2017) .

running variable is continuous. Continuous running variables are favorable, compared to discrete variables, for RD designs.

Figure 2.4c inspects the density around the birthdate cutoff, before the cutoff matters for scholarship receipt. There is a larger number of observations immediately to the right of the cutoff than immediately to the left. For 3 bins of 2 weeks to the right of the cutoff, there are at least 80 students and the 3 bins to the left have 75 or less students. Importantly this is not due to the scholarship, because the scholarship does not come into force until the next cohort. As the birthdates get farther from the cutoff, the increase in density levels out.

Figure 2.4d inspects the density around the birthdate cutoff, when the cutoff does matter for scholarship receipt. There is a similarly sized discontinuity in the distribution at the cutoff to Figure 2.4c when the cutoff does not matter for scholarship eligibility. For 3 bins to the right of the cutoff, there are about 75 students and for 3 bins to the left of the cutoff there are about 65 students.

As distance to the right of the cutoff increases, the difference in the distribution does not level out as much as Figure 2.4c. Since a goal of the law is to reduce state brain-drain (Sjoquist & Winters, 2014) and prior work finds an increase in total students (Scott-Clayton, 2011), this is not entirely unexpected. However, it is also not inconsistent with the identifying assumptions, which would require a discontinuous density right at the cutoff relative to the prior year.³⁸ One possible interpretation is that the scholarship had no effect on in-state attendance for those who are the oldest in their cohort.

Figures 2.4e and 2.4f present formal density tests of the distributions of number of observations around the cutoffs before it matters for scholarship receipt and when it does. The exact method for the implementation follows Cattaneo, Jansson, and Ma (2019). The histograms in Figures 2.4c and 2.4d required a choice for the bin width, but these density tests do not require pre-binning.³⁹

Figure 2.4e shows the test of the density at the cutoff, before the cutoff determines scholarship

³⁸There are 1,791 students to left and 1,792 students to the right of the cutoff in Figure 2.4c. There are 1792 to the left and 1861 to the right in Figure 2.4d.

³⁹Bin widths are selected based on asymptotic mean square error (AMSE).

receipt. The point estimate of the density to the left of the cutoff is .00144 and the point estimate on the right is .00185, a difference of 0.00041. The null hypothesis of smoothness is not rejected, with a p-value of 0.1618 which suggests there is no effect of relative age on enrolling at this university prior to the beginning of the scholarship.

Figure 2.4f repeats the same procedure for the cutoff that determines scholarship receipt. The point estimate left of the cutoff is 0.00133 and the point estimate to the right is 0.0014, a difference of 0.00007. The null hypothesis of smoothness is not rejected, with a p-value of 0.8131, which suggests there is no effect of the scholarship on enrolling, at the birth cutoff, at this university.⁴⁰ This test suggests there is no manipulation of birthday or selection into the sample at the cutoff.

2.4.3 Pre-Treatment Covariate Balance

One issue identified in the density test is that density no longer levels off after a few weeks past the cutoff, so this section investigates whether this results in students who are different along observable pre-treatment covariates close to either side of the cutoffs. If the average of pre-treatment covariates at the pre-treatment covariates is the same at the cutoff, then it also supports the identifying continuity assumptions. The standard method, that is used below, for detecting differences in pre-treatment covariates is to substitute the pre-treatment covariates for the outcomes of interest, then fit local linear regressions on both sides of the cutoff.

Scholarship and Pre-Scholarship Cutoffs

Table 2.1 presents results from pre-treatment covariate balance tests.⁴¹ Columns 1 and 4 present the results from the post-scholarship cutoff. Columns 2 and 5 present results from the pre-scholarship cutoff does not matter for scholarship receipt.

⁴⁰When attempting to impose strong assumptions on the higher order derivatives and cumulative distribution, which increase the power of the test, the optimal bandwidths exceed the limit of the data for the post-scholarship cutoff (Cattaneo, Jansson, & Ma, 2019).

⁴¹Figure B.6 shows plots that visualize these tests.

Scholarship Year In column 1, there are no statistically significant effects of being born just to the right of the cutoff on any outcome- male, ACT composite score, ACT English, or ACT Reading at the post-scholarship cutoff. Column 4 shows that there are no statistically significant effects of being born after the cutoff on SAT combined score, ACT Math, or ACT Science sections. Column 4, Panel A does show that students born after the cutoff in the year of the scholarship have a high school GPA that is 0.13 points higher than those to the left of the cutoff. This is a small effect, as it is only 1/3 of the difference between an A- and an A or an A and an A+ and only 1/2 to 1/4 of a standard deviation in high school GPA.

Pre-Scholarship In column 2, there are no statistically significant effects of being just beyond the cutoff, in the year before the cutoff, on ACT composite score (Panel B), ACT english score (Panel C), or ACT reading (Panel D). Panel A, column 2 does show that being just beyond the cutoff, in the year before the scholarship, shows a 14 pp decline in probability of being male. In column 5, there are no statistically significant effects of being beyond the cutoff on high school GPA (Panel A) or ACT math (Panel C). Those beyond the pre-scholarship cutoff have an 83.31 point increase in SAT Combined score, which is statistically significant at 95% confidence. However, their ACT science scores are 0.84 points lower, which is statistically significant at 90% confidence.

No Existence of Relative Age Effects on Pre-Treatment Covariates There is no consistent evidence that students on either side of the cutoff are different. Students treated by relative age at the pre-scholarship cutoff score higher on the SAT, but score lower on the ACT science scores. If the story about relative age leading to either more or less qualified students, it would be expected for the cutoff to affect all academic outcomes in the same direction.⁴²

⁴²Another possibility is that students who take the SAT and ACT are materially different in some unspecified way, although this seems unlikely.

Did Scholarship Affect the Difference at the Cutoffs?

Columns 3 and 6 contain the most informative results for assessing pre-treatment covariate balance. These columns show the difference between the two cutoffs. Because there is a possibility that these covariates are discontinuous due to relative age effects, before the scholarship, strong evidence of pre-treatment covariate imbalance requires that the scholarship cutoff widens this discontinuity beyond where it was previously. Sometimes, the estimate reported in columns 3 and 6 is not the exact difference of the preceding columns, because the bandwidth is constrained to a common number that also ensures there is no overlap of birthdates (and students) used in the estimation of the two discontinuities.

Of all estimates in column 3 and 6, the only one that is statistically significant is Panel A, column 3. The estimate in Panel A, column 3 estimates that the scholarship increased the probability of being male by 17 pp at the cutoff. While this is less than ideal, it is important to note that this is driven by a statistically significant and large coefficient on the pre-cutoff. ACT math and ACT science are nearly 10% significant, suggesting the post-cutoff increases pre-treatment academic ability relative to the the pre-cutoff. These effects are also driven primarily by the pre-cutoff.

Finally, there are no statistically significant differences between the pre and post cutoffs on either high school GPA (Panel A, column 6) or ACT composite score (Panel B, column 3). High school GPA is only 0.06 points higher at the post cutoff relative to the pre cutoff. The ACT composite score is only 0.47 points higher at the post scholarship cutoff compared to the pre scholarship cutoff.

It is not supported by the data that there are massive relative age related differences (columns 1, 2, 4, and 5) at the cutoffs. It is also not supported that there are large differences in college attendance decisions (i.e. highly qualified students choosing to attend in-state instead) at the cutoffs that are induced by the scholarship beginning (columns 3 and 6) . The pre-treatment covariate balance tests are thus consistent with the identifying assumption.

2.5 Methods

Fuzzy regression discontinuity is used to estimate the effect of scholarship receipt on academic performance. It exploits pseudo-random assignment of the scholarship, which is a result of state-wide kindergarten compulsory entry age assigning students to different cohorts based on small differences in birth-date. The local average treatment effect of the scholarship is estimated via a two-stage approach.

The first stage follows,

$$S_i = \lambda + \psi \text{After}_i + f(\text{Birthdate}_i) + \epsilon_i, \quad \forall | \text{Birthdate}_i | < h, \quad (2.7)$$

in which subscript i stands for individual and c stands for cohort. The effect of birthdate is a flexible local linear regression that, in both regressions, is fit separately on either side of the cutoff. S stands for scholarship receipt. After is a dummy variable that equals 1 if a student enters in cohort 2002, the first cohort that students were eligible. The coefficient ψ has been shown to be large, around .4 or .5 in Section 2.4.1.

To construct the Wald estimate of the local average treatment effect the reduced form is estimated by the following equation,

$$y_i = \alpha + \beta \text{After}_i + f(\text{Birthdate}_i) + e_i, \quad \forall | \text{Birthdate}_i | < h. \quad (2.8)$$

The estimated intent to treat (ITT) effect of the scholarship is β , the observed difference between the regressions on either side of the cutoff, at the cutoff. The Wald local average treatment effect of the scholarship is $\tau = \frac{\beta}{\psi}$.

Bandwidth and Kernel While the local linear regression approach above does not require specifying functional form, it does require selecting a bandwidth (h) and kernel for weighting. Using smaller bandwidths reduces bias and increases variance, so the mean-square error (MSE) optimal bandwidths are used, which optimally trades off bias and variance (Imbens & Kalyanaraman,

2012). A triangular kernel, which puts more weight on closer observations, is preferred, because it gives point estimates with optimal properties when used in tandem with MSE-optimal bandwidths (Cattaneo, Idrobo, & Titiunik, 2019, pg. 44).⁴³

2.5.1 Identification

Next, assumptions for τ to represent the causal effect of receiving the scholarship on academic performance are considered. First, the standard instrumental variables assumptions are described. Next, regression discontinuity assumptions are discussed.

First Stage, Independence, and Exclusion Equations 2.7 and 2.8 outline a fuzzy regression discontinuity approach, and since treatment does not go from 0 to 1 at the cutoff, the usual instrumental variable assumptions are required.⁴⁴ For the design to be valid, there must be a first stage, which is discontinuous change in the likelihood of scholarship receipt at the date cutoff. Section 2.4.1 establishes that this assumption is met. The exclusion restriction requires small differences in birthday affect outcomes only through scholarship receipt and independence requires that small differences in birthday are exogenous to scholarship receipt. As discussed in Section B.3, there are some concerns that these are violated; however, Sections 2.4.2 and 2.4.3 find no empirical evidence that this is the case.

Monotonicity Two other usual assumptions of instrumental variables are monotonicity and the Stable Treatment Unit Value Assumption (SUTVA.) Monotonicity simply means that being to the right of the date cutoff does not reduce the likelihood of entering in the later cohort and subsequently receiving the scholarship for any individual student which is reasonable. Furthermore, this builds on Scott-Clayton (2011), because the compliers that the estimated effects are valid for are students who finish one grade per year in this approach.

⁴³Results are similar when varying the kernel and bandwidth and also to using bias-correction procedures (Calonico, Cattaneo, & Titiunik, 2014).

⁴⁴Dong (2018) requires less assumptions.

Stable Unit Treatment Value Assumption Next, SUTVA assumes that the potential outcomes of students to the left of the cutoff are not affected by the treatment of students to the right of the cutoff. This assumption is quite reasonable for year 1 outcomes, but when students on either side of the cutoff are in school at the same time, one cohort being treated may cause students in the other cohort to also enroll in more credits to be on pace with peers. This is part of the reason to focus on the first year outcomes. Another reason to focus on 1 year outcomes is SUTVA assumes equal dosages of treatment, which is violated for long-term outcomes, because students lose their scholarship if they fall below the scholarship, as noted by [Carruthers and Özek \(2016\)](#).⁴⁵

Local Smoothness The identifying assumption is local smoothness of potential outcomes ([Dong, 2018](#); [J. Hahn et al., 2001](#)).⁴⁶ Under local smoothness of potential outcomes, it is valid to use units on either side of the cutoffs as counterfactuals for one another and calculate the treatment effect as the difference in outcomes at the cutoff.⁴⁷ Researchers support local smoothness using the McCrary Density test ([McCrary, 2008](#)) and ensuring that pre-treatment covariates are balanced for treatment and control groups. The tests presented in Sections [2.4.2](#) and [2.4.3](#) are consistent with local smoothness of potential outcomes.

2.6 Main Results

First, reduced form plots and estimates show the relationship between the date cutoff and academic outcomes. Then, the local linear regression discontinuity estimates are presented. Next, difference in discontinuity results confirm that the estimated effects are driven by scholarship receipt.

⁴⁵Figures [B.4](#) and [B.5](#) show that many lose their scholarship, but there is still a discontinuity when considering the treatment being 4 years of scholarship.

⁴⁶Figure [B.8](#) shows the identification assumption. To be a causal effect, potential outcomes must be smooth at the cutoff.

⁴⁷These potential outcomes are not observed, so this assumption is not directly testable. This is the essence of the “fundamental problem of causal inference,” ([Holland, 1986](#), pg. 947).

2.6.1 Reduced Form Effects

First, the reduced form effects are visualized and the intent to treat effects are visualized. The intent-to-treat effects are estimated first, because they are the numerator of the local average treatment effect. If the intent-to-treat effects are not zero at the post-scholarship cutoff for academically eligible students, are zero at the post-scholarship cutoff for academically ineligible, and are zero at the pre-scholarship cutoff for both types of students, then it is suggestive that the estimated effects are driven by scholarship receipt.

Reduced Form Plots

Figure 2.5 plots the reduced form relationships for academically eligible students at the pre and post-scholarship cutoffs. In Figures 2.5a and 2.5c, there is not obviously distinguishable effect of exceeding the pre-scholarship cutoff on outcomes. This suggests there is no detectable effect of being the oldest in one's cohort on the outcomes of interest.

At the post-scholarship cutoff shown in Figures 2.5b and 2.5d, there are more distinguishable effect of exceeding the cutoff on outcomes. In Figure 2.5b, there is an increase in from around 50% before the cutoff to near 70% after the cutoff. In Figure 2.5d, the mean before is around 60% and the mean after is around 70%. In Figures 2.5b and 2.5d the outcomes increase as the distance to the cutoff increases, which is consistent with how the scholarship receipt grows after the scholarship.⁴⁸

Intent to Treat Estimates

Table 2.2 presents ITT estimates, from Equation 2.8, across academically eligible and academically ineligible students for the pre and post scholarship cutoffs. Column 1, the estimated ITT effect of exceeding the post-scholarship cutoff for academically eligible students, is consistently estimated to be statistically significant at 90% confidence and in line with the approximate sizes seen in Figures 2.5a and 2.5c. Column 2 estimates the same effect, except at the pre-scholarship cutoff where relative age effects may be present but scholarship effects are not. Just as seen in Figures 2.5a and

⁴⁸This exercise is repeated for academically ineligible students in Figure B.9.

2.5c, there is no statistically significant effects of exceeding the pre-scholarship cutoff on exceeding the renewal credits or GPA thresholds. Column 3 shows the difference in the estimated ITT effects across the cutoffs, which is between 0.196 and 0.114, but often not statistically significant due to the increased imprecision that comes from estimating two separate discontinuities.⁴⁹

Column 4 repeats the estimation of ITT effects for academically ineligible students at the post-scholarship cutoff and there are no statistically significant estimates, which is consistent with the effects being driven by scholarship receipt and not relative age or time trends. Column 5 repeats the estimation of ITT effects for academically ineligible students at the the pre-scholarship cutoff and finds no statistically significant effects. Column 6 finds the difference at the pre and post scholarship cutoffs for academically ineligible students is between -0.048 and 0.104 and never statistically significant.⁵⁰ Figure 2.5 and Table 2.2 are strong support that estimates are not driven by relative age or season of birth confounding, they are driven by scholarship receipt primarily.

2.6.2 Local Average Treatment Effects from Fuzzy Regression Discontinuity

Table 2.3 presents estimates of scholarship receipt on first year academic outcomes. There are varying choices of kernel, sample, covariates, coefficient adjustments, and bandwidth selection criterion to account for several justifiable ways of estimating the effects. Table 2.3, Panel A focuses on the likelihood of completing 30 credits in ones' first year. In column 1, no bias-correction or robustness corrections to the estimates are done and the scholarship increases the likelihood of completing the scholarship by 36 percentage points, which is significant above 95 percent confidence. In column 2, the bias-correction and robustness adjustments are done (Calonico, Cattaneo, Farrell, & Titiunik, 2017), and the results are essentially unchanged.

In columns 1 and 2 (in Panel A and B) academically ineligible students are included, making the control group less ideal. In column 3, academically ineligible students are dropped and the

⁴⁹Column 3 restricts the bandwidths, h , to 160 to ensure that no observations are used in the estimation of both discontinuities.

⁵⁰Also, the positive estimate of the difference are driven by a large negative at the pre-scholarship cutoff for academically ineligible students.

effect size decreases to 27 percentage points, probably due to a more academically gifted control group being utilized. Additionally, the statistical significance drops to 90%. In column 4, academic controls for HS GPA and entry test (ACT/SAT) scores are included to compare individuals of similar scores.⁵¹ The effect size drops to 0.24, which is the exact point estimate [Scott-Clayton \(2011\)](#) reports, but the coefficient is no longer statistically significant at conventional levels ($p < 0.128$). Column 5 uses CER-optimal bandwidths, which can be more desirable for inference, but less desirable for point estimates. The estimated effect does not change, but the standard error grows, reducing the p-value.

Columns 6 and 7 use a uniform kernel, because it minimizes asymptotic variance under certain conditions ([Cattaneo, Idrobo, & Titiunik, 2019](#), pg. 44). The difference in scholarship receipt at the cutoff increases noticeably with the distance from the cutoff for a few months before flattening, so triangular kernels place more weight on potentially unaffected observations. Column 6 uses the MSE-optimal bandwidth criterion with a uniform kernel, the effect size remains similar, but the standard error drops and the effect is statistically significant above 90% ($p < 0.085$). In column 7, a uniform kernel is used with the coverage error rate optimal bandwidths, the effect size grows to a 31 percentage point increase. The standard error also grows, causing this estimate to remain statistically significant at 94.6 percent confidence. The point estimates are similar to [Scott-Clayton \(2011\)](#), sometimes slightly larger depending on bandwidth, sample, controls, and correction.

[Table 2.3](#) Panel B replicates Panel A, using the likelihood of meeting the renewal GPA of 2.75 as the dependent variable. Column 1 shows the scholarship increases the likelihood of hitting the GPA threshold for renewal by 31 percentage points. The effect is statistically significant above 95%. Columns 2-4 add parameter corrections, subsample to academically eligible students, and include academic controls; all changes leave the effect size and statistical significance almost unchanged. Column 5 uses the CER-optimal bandwidth and finds a 33% larger effect. This point estimate should be interpreted cautiously as the point estimate does not have optimal point estimate

⁵¹Non-academically eligible students are dropped before academic controls are included, because covariate inclusion in RD assumes that the controls are additively separable from the running variable ([Calonico, Cattaneo, Farrell, & Titiunik, 2019](#)). Dropping non-academically eligible students increases the plausibility of this assumption.

properties (Cattaneo, Idrobo, & Titiunik, 2019, pg. 44).

Columns 6 and 7 of Table 2.3, Panel B use the uniform kernel. In column 6, the coefficient is 33% smaller, but the standard error is reduced. Still, the coefficient is only statistically significant at 9% confidence. In column 7, a uniform kernel is used with the CER-optimal bandwidths and the estimate becomes significant at 95 percent confidence. and the effect grows back to near the previously estimated effects. These estimates are similar to the increased likelihood of completing the renewal credits that are presented in Panel A of Table 2.3, in contrast to Scott-Clayton (2011).^{52,53}

[DISCUSS THE CONTINUOUS RESULTS AND WHY HIGHER CONTINUOUS GPA RESULTS ARE DRIVING LARGER GPA RENEWAL ESTIMATE]

2.6.3 Difference in Discontinuity Estimates

Second, difference in discontinuity is used, which takes advantage of the cutoff being the exact same date in the prior year. Under certain assumptions relative age can be differenced out of the treatment effect estimate. To address the confounding effect of differential relative age at the entry date cutoff is difference-in-discontinuity is used (Canaan, 2020; Grembi et al., 2016; Smith, 2016).⁵⁴ The effect of relative age at the cutoff is present in the years before the scholarship becomes available. To remove the confounding effect of relative age, one simply subtracts off the effect at the birth cutoff in prior periods

$$T^{DD} \equiv (Y^- - Y^+) - (\tilde{Y}^- - \tilde{Y}^+),$$

where \tilde{Y} are the outcomes at the cutoff in period before the cutoff determines scholarship receipt. To implement this estimator, the data is formatted so that 6 months on either side of each cutoff are related to that particular cutoff. Effectively, the difference in discontinuity estimate is the local aver-

⁵²Table B.3 estimates parametric functions that assume functional forms for the birthdate running variable and find similar results when dropping birthdates farther away from the cutoff.

⁵³The continuous results are shown in Table B.2.

⁵⁴This estimator differs from other RD estimators that exploits longitudinal data, because the variation it uses is cross-sectional, not within-unit. Other estimators include fixed effect RD estimators (Pettersson-Lidbom, 2012), first-difference RD estimator (Lemieux & Milligan, 2008), and dynamic RD designs (Cellini, Ferreira, & Rothstein, 2010).

age treatment effect (τ , shown in [Table 2.3](#)) minus the ITT effect ($\beta_{\text{pre-cutoff}}$), at the pre-scholarship cutoff, from [Equation 2.8](#) (shown in [Table 2.2](#), column 2).

Identification Assumptions

A new assumption is the effect of relative age, in the case of no treatment, is constant over time. This is similar to the parallel trends assumption for differences-in-differences, with the important difference that this assumption is local. Under typical IV assumptions (due to the fuzzy cutoff), local smoothness, and local “parallel trends”, T^{DD} is a causal effect of receiving the scholarship.

Assumption to Interpret T^{DD} as Local Average Treatment Effect

One more assumption is required for T^{DD} to represent the local causal effect of receiving the scholarship. The final assumption is that the effect of receiving the scholarship does not depend on relative age. Stated differently, there must be no interaction between relative age effects and the effects of the scholarship. There must be additive separability between relative age and receiving the scholarship.

Results

[Figure 2.6](#) presents difference in discontinuity estimates of the effect of the scholarship. [Figure 2.6a](#) plots the difference in discontinuity estimates on credits for different bandwidths and kernels. For all combinations of bandwidth and kernel the point estimates of receiving the scholarship, due to exceeding the birthdate cutoff and netting out potential relative age effects, is 0.24. Most of the estimates are statistically significant at 90% confidence. This is nearly the exact same estimate for credits as [Scott-Clayton \(2011\)](#).

[Figure 2.6b](#) plots difference in discontinuity estimates on GPA for different bandwidths and kernels. For most combinations of kernels and bandwidths, the estimated effects are much closer to 0.24 than 0.08, as in [Scott-Clayton \(2011\)](#). Most estimates are also statistically significant at 90% confidence. The only kernel and bandwidth that comes close to estimating the 0.08 pp effect

that is seen in [Scott-Clayton \(2011\)](#) is the uniform kernel using students that are 160 days away from the cutoff on both sides. This is the evidence that suggests the large difference previously estimated is due to functional form limitations of parametric 2SLS, where the instrument is solely cohort of entry.

2.6.4 Heterogeneous Results

Finally, there may be heterogeneous effects by subgroups. One subgroup analysis is by academic credentials, because using the ACT cutoff for an RD does not permit estimating effects for those near the top of the academic ability distribution. The second subgroup analysis is by gender, because there are gender gaps in education.

By Academic Credentials [Figure 2.7](#) investigates effects for academically eligible students who are above and below the median ACT (24) and HS GPA (3.71) (for academically eligible students). The estimates are noisy, due to small sample sizes; however, some differences in point estimates do arise by above/below median academic scores. First, in [Figure 2.7a](#), there is a larger point estimate for exceeding credits for those who are above median ACT than those below (or equal to) median ACT. Second, in [Figure 2.7d](#), there is a larger point estimate for exceeding GPA for those who are below the median high school GPA. These are important results for beginning to untangle the heterogeneity of program effects across those with different academic credentials.

By Gender Finally, [Figure 2.8](#) shows difference in discontinuity effects by gender. The outcomes are binary for whether or not the students finished the GPA or credits renewal requirement in each of their first 3 years. [Figure 2.8a](#) shows there is no difference between males and females of finishing the credits requirement for 3 years. [Figure 2.8b](#) shows males are more likely to complete the GPA requirement for 3 years, but women are not.⁵⁵ These estimates are important in the context of the known gender gaps in education.

⁵⁵Figures [B.10](#) and [B.11](#) shows estimates by year by gender.

2.7 Conclusion

Merit-aid programs have grown significantly recently; however, they may be poorly targeted and compete with other programs for scarce state funding. It is difficult to estimate the causal effects of merit-aid, because students have an incentive to sort themselves into treatment, whether it be through retaking academic scores or by attending in-state when they would not have without merit-aid. Prior solutions to estimating causal effects are not applicable to West Virginia, because the academic cutoffs are known and no data has been able to distinguish between first and best scores.

This paper proposes, formalizes, and validates an alternative empirical strategy that provides a way to estimate causal effects of merit-aid which is compatible with a wider range of datasets. This alternative variation compares students born on slightly different days, 19 years before the scholarship begins, who have vastly different probabilities of receiving the scholarship due to Kindergarten cohort cutoff dates and regular grade progression. Students are not more likely to sort into treatment at this cutoff, unlike academic cutoffs, and are similar along observable pre-treatment characteristics. Furthermore, there are no relative age effects that are detectable at this cutoff.

Receiving a merit-aid scholarship leads to an equally high likelihood, 24 percentage points, of completing both the credits and GPA renewal requirements unlike prior work that is constrained by functional form assumptions ([Scott-Clayton, 2011](#)). The scholarship bolsters the credits completed by those above the median ACT score more than those below it and bolsters the GPA performance of those with lower high school GPAs. Finally, the scholarship leads to men completing the GPA requirements for 3 years, but not women.

There are two primary limitations. The first is that the dataset does not contain all students in West Virginia, so it is possible that differences are driven by sample composition. The second is that there are relatively few observations observed at the cutoff, so statistical power is limited, especially in sub-group analyses. Both limitations are addressable with additional data from other universities in West Virginia and in other merit-aid states that have known academic cutoffs.

This research has significant policy implications. Academic renewal thresholds were found to be ineffective in Georgia ([C. M. Cornwell et al., 2005](#)) and effective in West Virginia ([Scott-](#)

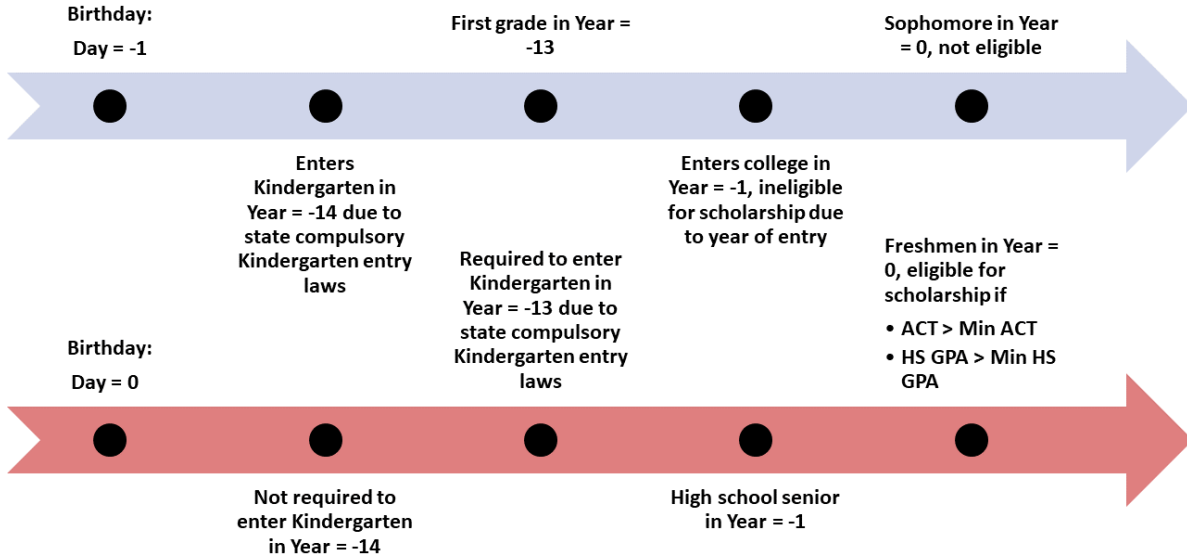
Clayton, 2011). This research finds that both credits and GPA renewal thresholds are equally effective, when there is a limit on the number of semesters that the aid can be used. Other financial incentives for academic performance can use defined thresholds as a way to increase credits completed and GPA.

There are a number of extensions for future research. First, as mentioned above, is this similarity across types of academic performance the same for other college students in the same state? Second, since the cutoff adds both a goal and financial incentive, is there a way to separately identify the behavioral impact of the goal without financial aid? Third, knowing the difference between c and c' could help identify how much attrition from the expected grade path occurs. This would add to the knowledge about who the compliers are. Fourth, is this same research design applicable to other programs and/or other states? If so, this research design could help advance the knowledge of causal impacts of education programs.

Figure 2.1: Illustration of Scholarship Differentially Assigned by Small Changes in Birth Date, Compulsory Kindergarten Entry Date Cutoff Laws, and Grade Progression

(a): Kindergarten Cutoff that Matters for Scholarship Receipt

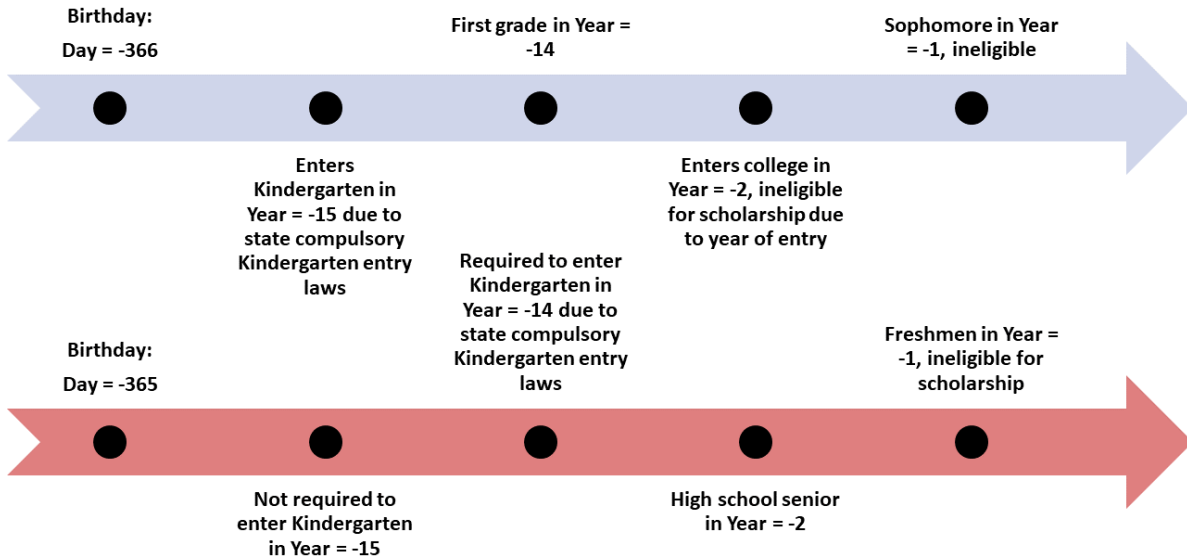
Control Group 1



Treated by Scholarship and Relative Age

(b): Kindergarten Cutoff that Doesn't Matter for Scholarship Receipt

Control Group 2

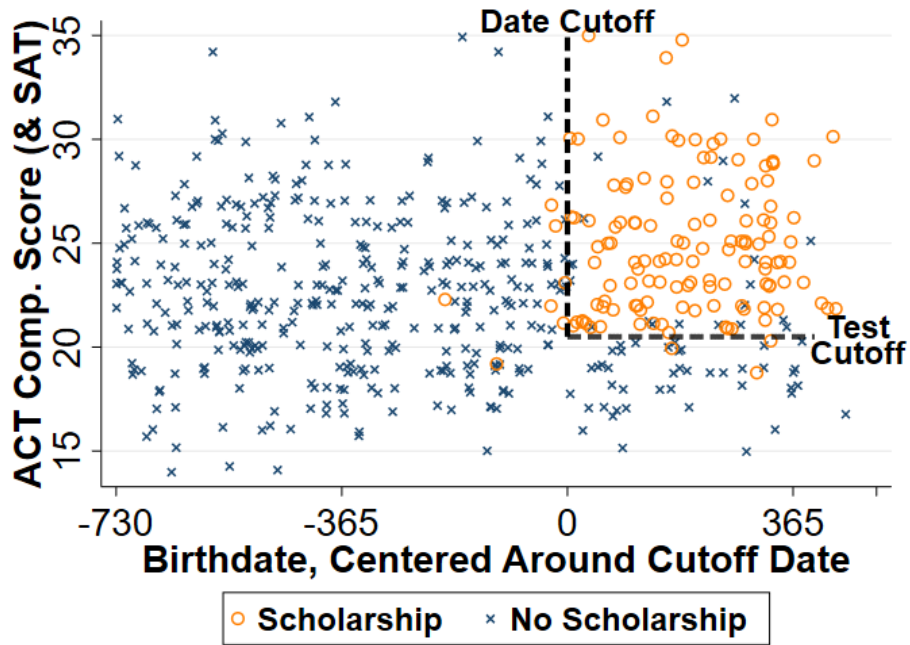


Treated by Relative Age

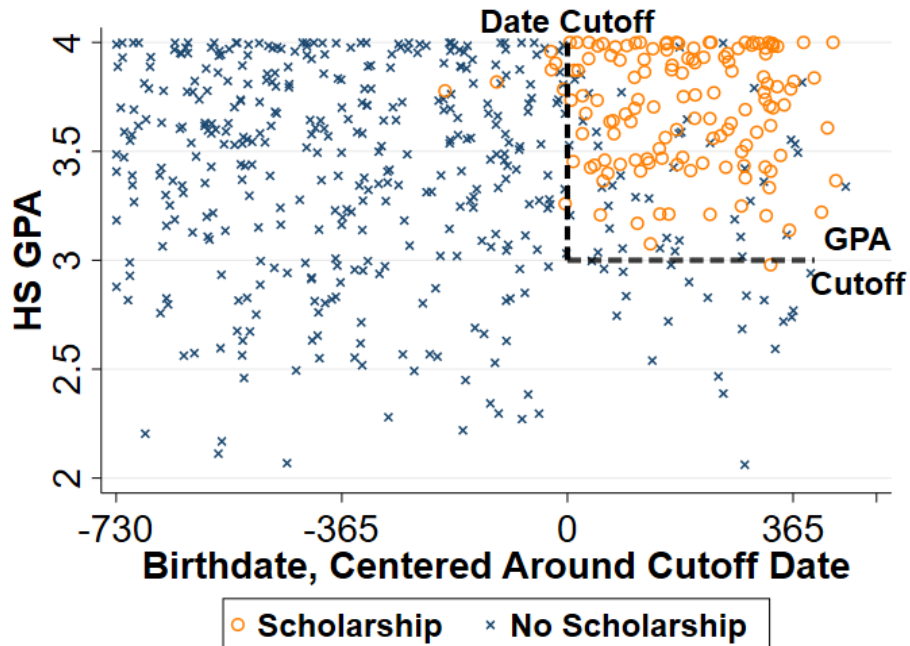
Note: Year 0 stands for initial year of this college scholarship. Day = -1 is a student born right before the birthdate cutoff to enter kindergarten in the first cohort that is eligible for the college scholarship if they finish 1 grade per year. Day = 0 is a student born right after the cutoff which allows them to enter in the kindergarten cohort that will enter college in the first college eligible cohort, finishing 1 grade per year.

Figure 2.2: Eligibility Measures and Scholarship Receipt

(a): Birthdate, Entry Test Score, and Scholarship



(b): Birthdate, HS GPA, and Scholarship

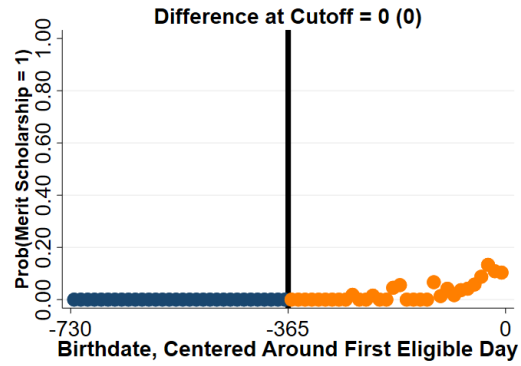
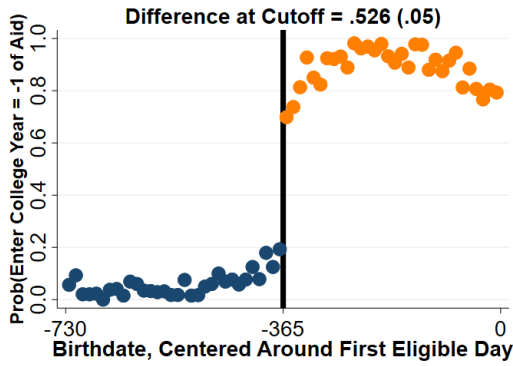


Note: 10% random sample. N = 559. Birthdates centered at 0, the first eligible date for the scholarship. The cutoff ACT score is 20.5. SAT score converted to ACT using the conversion chart at [Edwards \(2020\)](#). The cutoff HS GPA is 3.0.

Figure 2.3: Kindergarten Entry Date Cutoffs on Cohort Entry Year and Scholarship Receipt

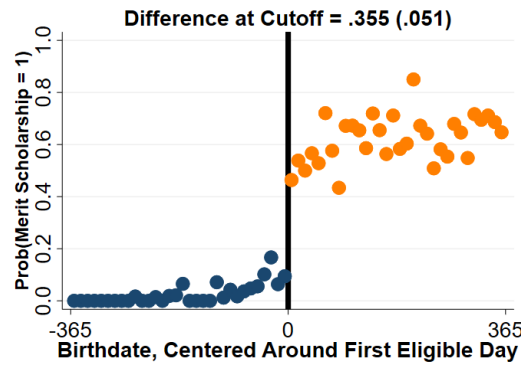
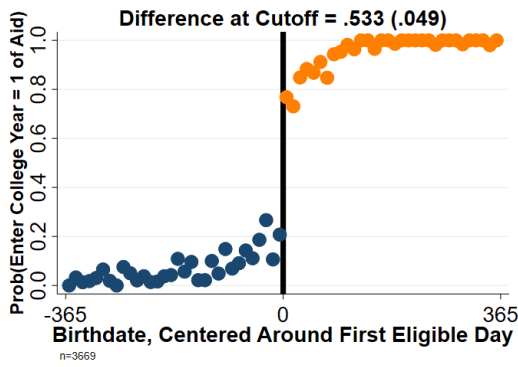
(a): Pre-Scholarship Cutoff on Enter in 2001

(b): Pre-Scholarship Cutoff on Scholarship Receipt



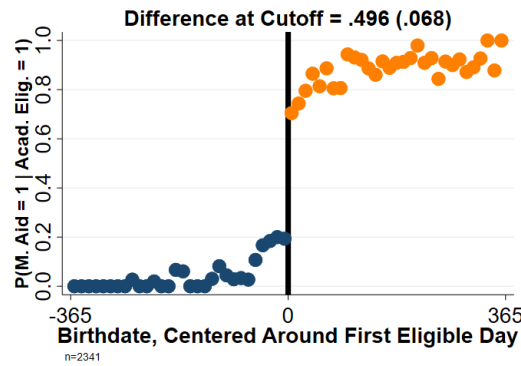
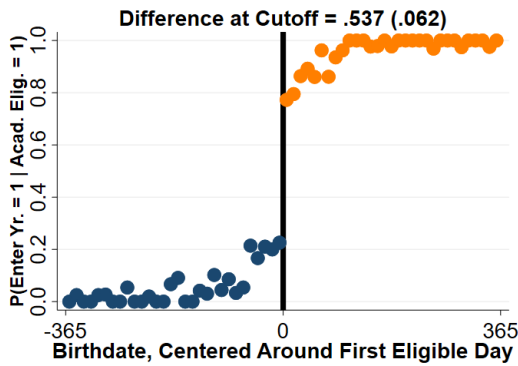
(c): Post-Scholarship Cutoff on Enter in 2002

(d): Post-Scholarship Cutoff on Scholarship Receipt



(e): Post-Scholarship Cutoff on Enter in 2002, Academically Eligible Students

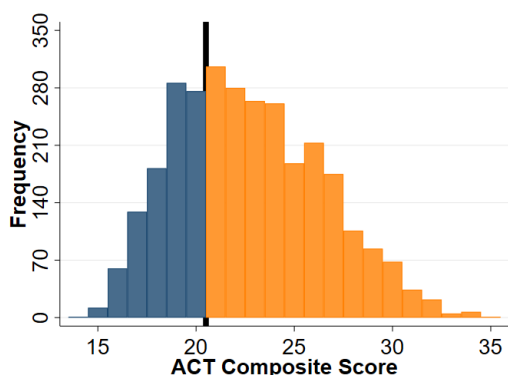
(f): Post-Scholarship Cutoff on Scholarship Received, Academically Eligible Students



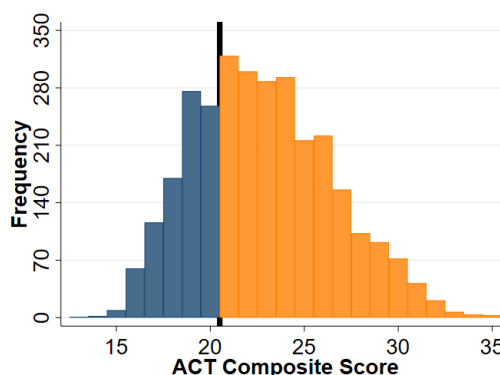
Note: Plots show the conditional mean within birthdate bins. Bin-length = 11.375 days. The black vertical line indicates cutoff birthdates. The birthdate is centered at the post-scholarship cutoff birthdate, meaning 0 is the first birthdate of increased likelihood of scholarship treatment. In Panels A, C, and E, the y-variable is 1/0 for being a first time freshman in the first year of the scholarship. In Panels B, D, and F, the y-variable is mean of receiving at least 1 year of financial merit-based aid. In Panels A and B, $n = 3,606$. In Panels C and D, $n = 3,669$. In Panels E and F, $n = 2,341$. The difference at the cutoff is estimated using local linear regression with a triangular kernel and a two-sided mean-square error (MSE) optimal bandwidth.

Figure 2.4: Comparing Academic and Birthdate Running Variables' Continuity Around Cutoffs

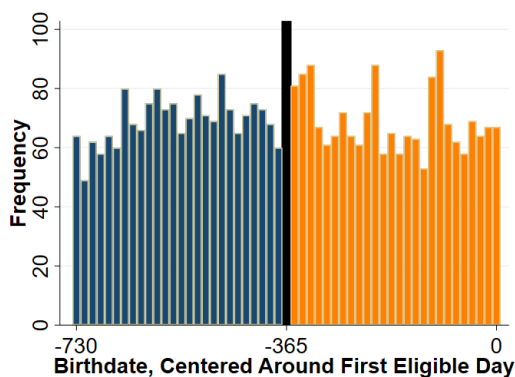
(a): ACT Scores, Two Cohorts Pre-Scholarship



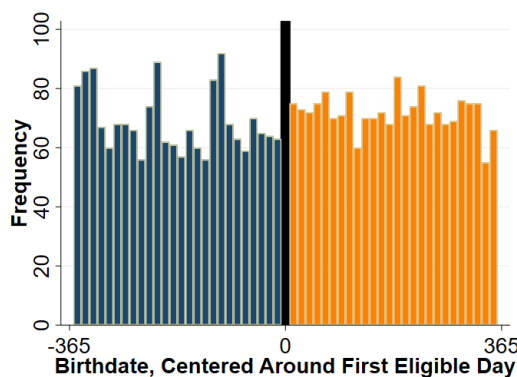
(b): ACT Scores, First Cohort of Scholarship



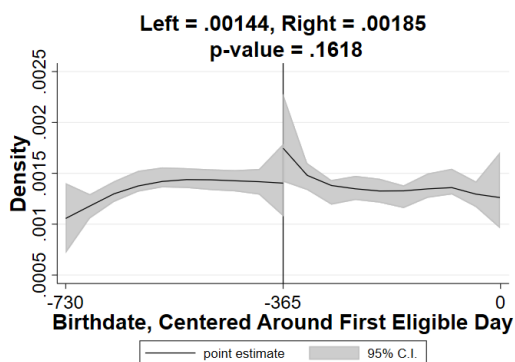
(c): Birthdate Density Around Cutoff, Before Scholarship



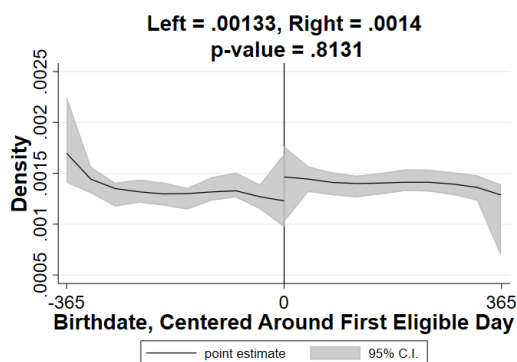
(d): Birthdate Density Around Cutoff, For Scholarship



(e): McCrary Density Test, Pre-Scholarship



(f): McCrary Density Test, First Scholarship Cohort



Note: Panels A and B show the frequency of each ACT Composite score realization. Panel A uses the 2000 and 2001 cohorts. Panel B uses the 2000 cohort. Birthdate 0 is the cutoff for the post-scholarship cutoff and birthdate -365 is the pre-scholarship cutoff. Panel C ($n = 3,600$) shows the distribution of those born within 1 year of the pre-scholarship cutoff date. Panel D ($n = 3,666$) shows the distribution of birthdate for those born within a year of the post-scholarship cutoff date. Each bar in Panels C and D represent 2 weeks of birthdates. Panels E and F are density tests at the pre and post-scholarship birthdate cutoffs. The estimates are calculated according to [Cattaneo, Jansson, and Ma \(2019\)](#). Left is the point estimate to the right of the cutoff and right is point estimate to the right, both are bias-corrected. The p-value is from a test for whether density to the left equals the density to the right and is robust. Failure to reject indicates that the density is smooth across the cutoff, consistent with identifying assumptions ([McCrary, 2008](#)).

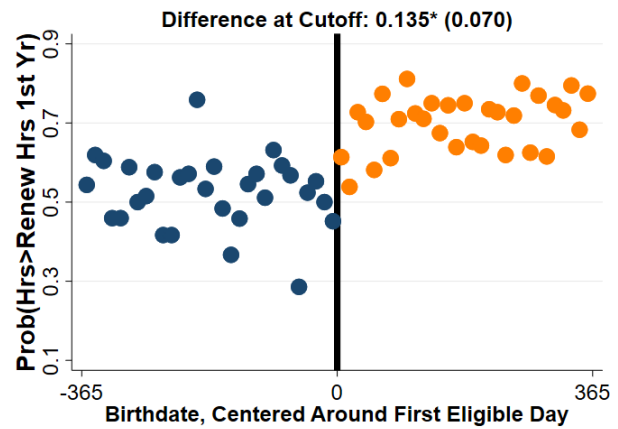
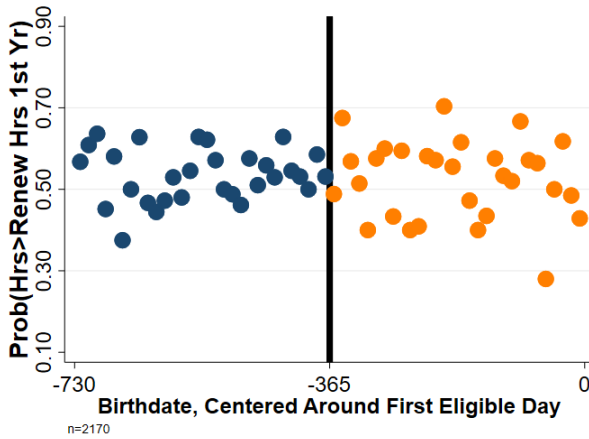
Table 2.1: Pre-Treatment Covariate Balance

Panel A						
	Dep. Var.: Male			Dep. Var.: HS GPA		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.03 (0.06)	-0.14** (0.06)	0.17** (0.08)	0.13** (0.05)	0.07 (0.05)	0.06 (0.07)
Bandwidth	149	154	150	189	166	160
Observations	1489	1569	3038	1861	1669	3231
Panel B						
	Dep. Var.: ACT Composite			Dep. Var.: SAT Combined		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	-0.05 (0.44)	-0.44 (0.44)	0.47 (0.62)	40.59 (33.68)	83.31** (37.06)	-42.76 (53.54)
Bandwidth	167	156	160	190	180	160
Observations	1362	1299	2661	337	319	591
Panel C						
	Dep. Var.: ACT English			Dep. Var.: ACT Math		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.01 (0.55)	0.05 (0.53)	-0.02 (0.73)	0.11 (0.48)	-0.73 (0.52)	0.90 (0.77)
Bandwidth	146	146	160	198	173	160
Observations	1199	1199	2661	1616	1427	2661
Panel D						
	Dep. Var.: ACT Reading			Dep. Var.: ACT Science		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	-0.31 (0.60)	-0.35 (0.58)	0.09 (0.84)	0.00 (0.41)	-0.84* (0.49)	0.91 (0.63)
Bandwidth	179	153	160	207	137	160
Observations	1446	1282	2661	1683	1132	2661

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Coefficients are from local linear regressions with triangular kernels and measure the discontinuity at the cutoff. The coefficients are not bias-corrected and the SE's are not robust. Mean-square error optimal bandwidths are used. The post-scholarship cutoff ($d=0$) cutoff is the cutoff that determines scholarship eligibility. The pre-scholarship ($d = -365$) cutoff is the cutoff on the same day in the prior year and does not determine scholarship eligibility. Difference may be slightly different than the difference of the two coefficients, due to different bandwidth lengths. Bandwidth is two-sided, meaning that many observations on both sides of the cutoff.

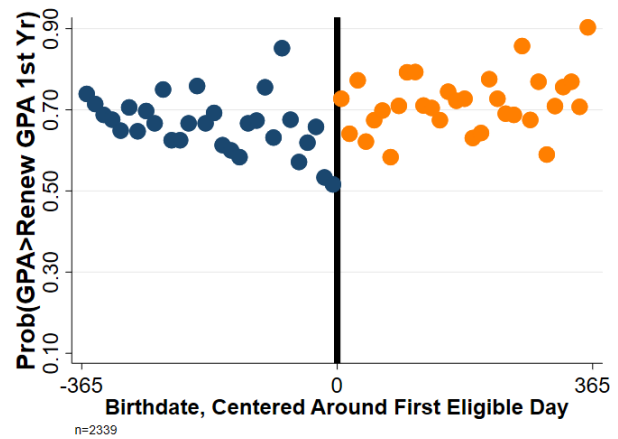
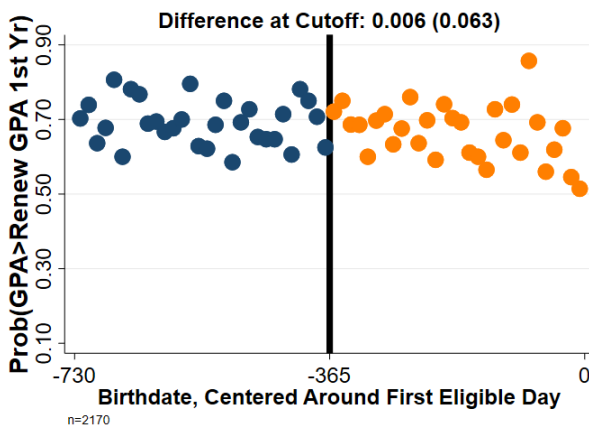
Figure 2.5: Effect of Being Born After Birthdate Cutoffs on Likelihood of Exceeding Renewal Thresholds in Freshman Year, Academically Eligible Students Only

(a): Exceed Renewal Credits, Pre-Scholarship Cutoff (b): Exceed Renewal Credits, Scholarship Cutoff



(c): Exceed Renewal GPA, Pre-Scholarship Cutoff

(d): Exceed Renewal GPA, Scholarship Cutoff



Note: Plots show the conditional mean within birthdate bins. Bin-length = 11.375 days. The black vertical line represents cutoff birthdates. The birthdate is centered at the first birthdate where increased scholarship receipt is expected, the post-scholarship cutoff is day = 0. Panel A shows the effect of being born later than the pre-scholarship cutoff on the likelihood of exceeding the freshman credits renewal threshold (≥ 30) in one's freshman year. Panel B shows the effect of being born later than the post-scholarship cutoff on the likelihood of exceeding the credits renewal threshold in one's freshman year. Panel C shows the effect of being born later than the pre-scholarship cutoff on exceeding the freshman year GPA renewal threshold (2.75). Panel D shows the effect of being born later than the post-scholarship cutoff on exceeding the freshman year GPA renewal threshold.

Table 2.2: Intent to Treat Effect on Meeting Academic Renewal Threshold By Academic Eligibility and Date Cutoff

Panel A: Credits, Conventional						
	Academically Eligible			Academically Ineligible		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.135*	0.011	0.129	0.055	-0.070	0.104
	(0.070)	(0.070)	(0.097)	(0.076)	(0.076)	(0.102)
Bandwidth	161	153	160	168	128	160
Observations	1021	926	1995	599	512	1236
Panel B: Credits, Bias-Corrected and Robust SE's						
	Academically Eligible			Academically Ineligible		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.122	0.019	0.114	0.051	-0.081	0.087
	(0.084)	(0.084)	(0.148)	(0.091)	(0.089)	(0.143)
Bandwidth	161	153	160	168	128	160
Observations	1021	926	1995	599	512	1236
Panel C: GPA, Conventional						
	Academically Eligible			Academically Ineligible		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.100*	0.006	0.127	0.049	0.036	0.001
	(0.057)	(0.063)	(0.092)	(0.085)	(0.084)	(0.120)
Bandwidth	231	151	160	180	141	160
Observations	1490	909	1995	629	564	1236
Panel D: GPA, Bias-Corrected and Robust SE's						
	Academically Eligible			Academically Ineligible		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.116*	0.006	0.196	0.038	0.064	-0.048
	(0.067)	(0.075)	(0.133)	(0.102)	(0.096)	(0.170)
Bandwidth	231	151	160	180	141	160
Observations	1490	909	1995	629	564	1236

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Estimates are the reduced form relationship for being born after the cutoff. Column 1 and 4 shows the estimates at the post-scholarship cutoff, column 2 and 5 shows estimates at the pre-scholarship cutoff, column 3 and 6 show the estimated differences. Columns 1, 2, 4, and 5 use mean square error optimal bandwidths. Columns 3 and 6 use bandwidths of 160 days at both cutoffs.

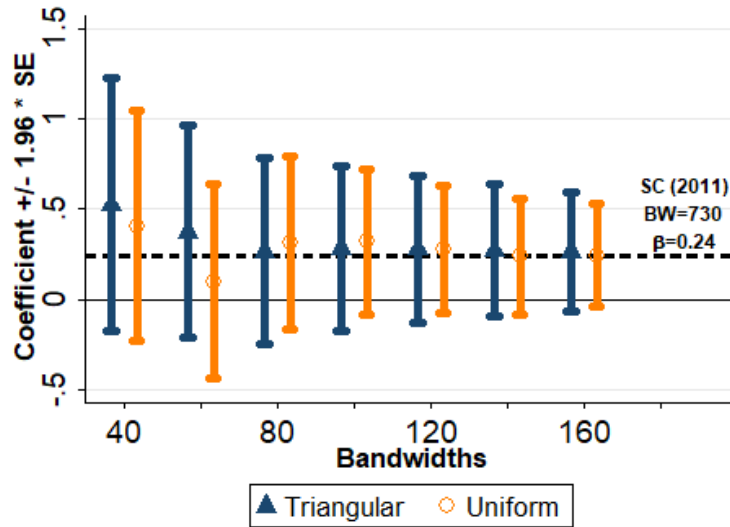
Table 2.3: Local Linear Estimates of Scholarship on First Year Credits and GPA

Panel A: Dep. Var.: Renew Credits							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Scholarship	0.36**	0.35**	0.27*	0.24	0.24	0.23*	0.31*
	(0.15)	(0.18)	(0.16)	(0.16)	(0.19)	(0.13)	(0.16)
Corrections	-	B-C, R	B-C, R	B-C, R	B-C, R	B-C, R	B-C, R
Sample	All	All	Ac. Elg.	Ac. Elg.	Ac. Elg.	Ac. Elg.	Ac. Elg.
Bandwidth	141	141	135	134	91	133	90
Eff. Obs. Left	680	680	395	395	248	393	248
Eff. Obs. Right	740	740	471	471	319	462	319
Obs-Left	1970	1970	1081	1081	1081	1081	1081
Obs-Right	1979	1979	1356	1356	1356	1356	1356
Acad. Controls				X	X	X	X
BW Criterion	MSE	MSE	MSE	MSE	CER	MSE	CER
Kernel	Tri	Tri	Tri	Tri	Tri	Uni	Uni
Panel B: Dep. Var.: Renew GPA							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Scholarship	0.31**	0.34**	0.34**	0.30**	0.40**	0.20	0.32**
	(0.15)	(0.18)	(0.16)	(0.15)	(0.19)	(0.13)	(0.16)
Corrections	-	B-C, R	B-C, R	B-C, R	B-C, R	B-C, R	B-C, R
Sample	All	All	Ac. Elg.	Ac. Elg.	Ac. Elg.	Ac. Elg.	Ac. Elg.
Bandwidth	151	151	140	142	96	138	93
Eff. Obs. Left	720	720	404	408	268	398	261
Eff. Obs. Right	777	777	491	500	336	482	327
Obs-Left	1970	1970	1081	1081	1081	1081	1081
Obs-Right	1979	1979	1356	1356	1356	1356	1356
Acad. Controls				X	X	X	X
BW Criterion	MSE	MSE	MSE	MSE	CER	MSE	CER
Kernel	Tri	Tri	Tri	Tri	Tri	Uni	Uni

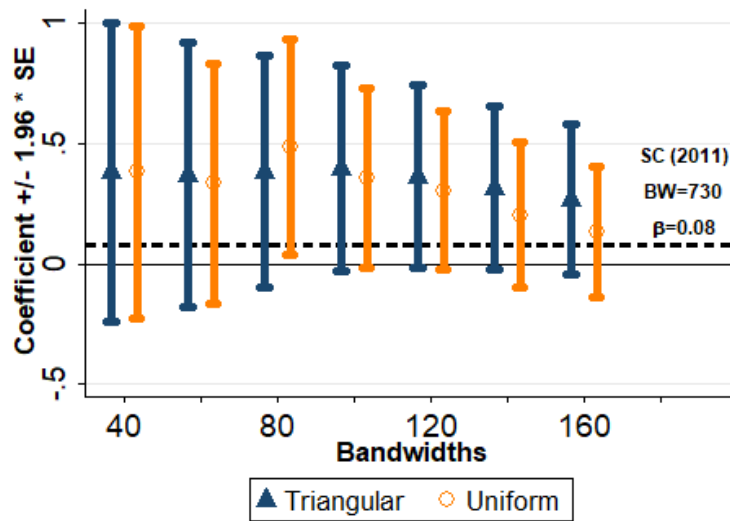
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Estimates displayed are the second stage Wald estimator from a fuzzy regression discontinuity design using birthdate in relation to Kindergarten entry date as the cutoff/instrument. The estimate is the effect of receiving the Promise scholarship. In all panels, the first stage cutoff has a strong effect on scholarship receipt. In all results, bandwidths are calculated using [CITE FOR MSE-OPTIMAL BW'S], in which a common bandwidth is assumed on each side. In columns 1 and 2, the full sample is used. In column 3, only students who exceed the academic cutoffs are included in the estimation sample. In columns 2 and 3, the bias-correction and standard error adjustment in [Calonico et al. \(2017\)](#) are used, because of the additional bias from using MSE-optimal bandwidths. Panel A the dependent variable is a dummy for finishes 30 credits (the renewal requirement) freshman year. Panel B the dependent variable is the number of continuous credits completed freshman year. Panel C the dependent variable is a dummy for exceeds a 2.75 GPA (the renewal requirement) freshman year. In Panel D the dependent variable is continuous GPA.

Figure 2.6: Coefficient Estimates from Difference in Discontinuity on Freshman Outcomes

(a): Probability of Completing Renewal Credits



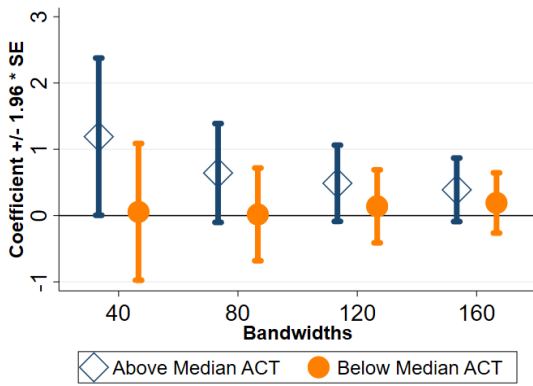
(b): Probability of Completing Renewal GPA



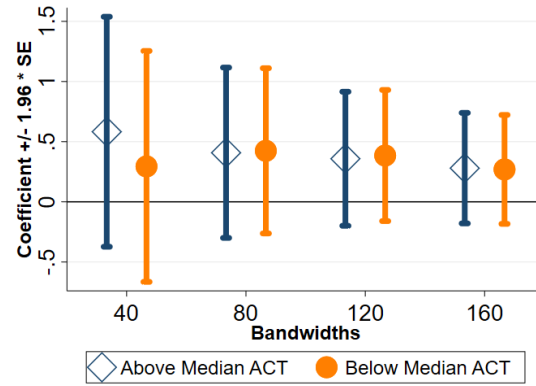
Note: Coefficients represent “conventional”, not bias-corrected or robust for MSE-optimal, treatment estimates from difference-in-discontinuity estimates (Grembi et al., 2016). Diff-in-discontinuity is the RD estimate from the kindergarten entry date cutoff before it determined the scholarship receipt, but did determine relative age subtracted from the RD estimate in where it did determine scholarship receipt. This approach differences out the relative age effect under fairly mild assumptions. The different bandwidths indicate the number of days used from the respective cutoffs. Triangular and uniforms kernels used. Panel B legend valid for both. All regressions subsample to academically eligible students meaning [(ACT \geq 21 OR SAT \geq 1000) AND HS GPA 3.0] and include controls for HS GPA and ACT composite score (w/ SAT converted to ACT). Panel A dependent variable is exceed renewal credits in freshman year. Panel B dependent variable is renewal GPA in freshman year. The horizontal dashed lines represent the Scott-Clayton (2011) point estimates which correspond approximately to 730 day bandwidths, since Scott-Clayton (2011) uses 2 cohorts before and after.

Figure 2.7: Heterogeneity by Academic Credentials

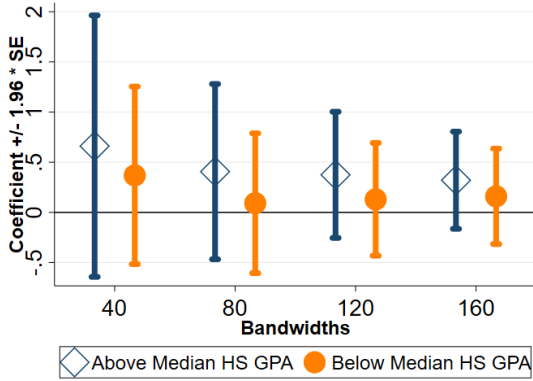
(a): Credits by ACT



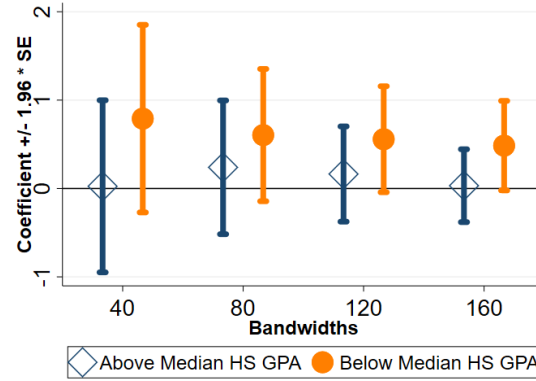
(b): GPA by ACT



(c): Credits by HS GPA

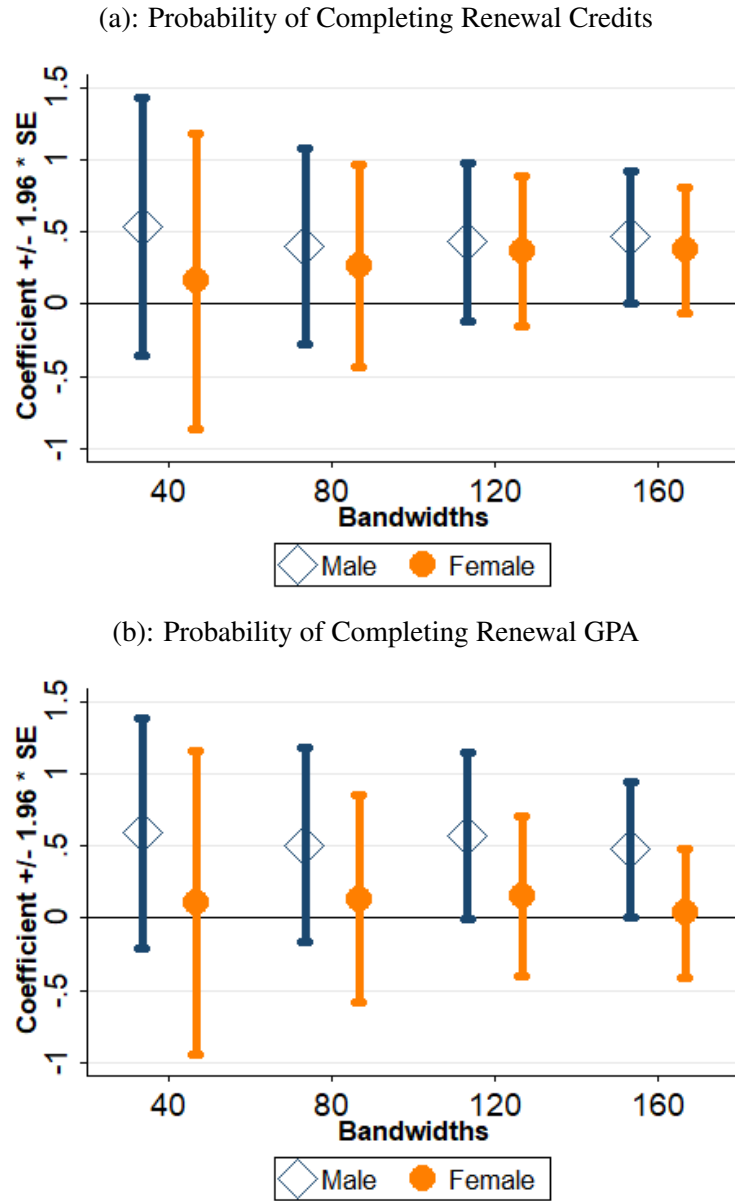


(d): GPA by HS GPA



Note: Difference in discontinuity estimates at different bandwidths plotted with 95% confidence intervals. The estimates are conventional. The local linear regression uses a triangular kernel. In Panels A and C, the dependent variable is exceeding the freshman credits renewal threshold. In Panels B and D, the dependent variable is exceeding the freshman GPA renewal threshold. The median HS GPA in this sample, for those who are academically eligible, is 3.71. The median ACT score in this sample is 24. The below group contains the median in all figures.

Figure 2.8: Coefficient Estimates from Difference in Discontinuity on 3 Year Outcomes, by Gender



Note: Coefficients represent “conventional”, not bias-corrected or robust for MSE-optimal, treatment estimates from difference-in-discontinuity estimates (Grembi et al., 2016). Diff-in-discontinuity is the RD estimate from the kindergarten entry date cutoff before it determined the scholarship receipt, but did determine relative age subtracted from the RD estimate in where it did determine scholarship receipt. This approach differences out the relative age effect under fairly mild assumptions. The different bandwidths indicate the number of days used from the respective cutoffs. Triangular and uniforms kernels used. Panel B legend valid for both. All regressions subsample to academically eligible students meaning [(ACT 20.5 OR SAT \geq 1000) AND HS GPA 3.0] and include controls for HS GPA and ACT composite score (w/ SAT converted to ACT). Panel A dependent variable is exceed renewal credits in freshman year. Panel B dependent variable is renewal GPA in freshman year. The horizontal dashed lines represent the Scott-Clayton (2011) point estimates which correspond approximately to 730 day bandwidths, since Scott-Clayton (2011) uses 2 cohorts before and after.

Chapter 3

The Effects of Random Peers on Noncognitive Abilities: Evidence from the National Outdoor Leadership School

In a survey of 3,000 employers, in which 1 was least important and 5 was most important, employers ranked academic performance and scores on tests a 2.5. The same employers ranked attitude a 4.6 and communication a 4.2 ([Bureau of the Census, 1998](#)), demonstrating employers believe noncognitive abilities are of utmost importance. The value employers place on non-cognitive skills has grown substantially and will likely continue to do so ([D. J. Deming, 2017](#)) and social skills are becoming more complementary ([Weinberger, 2014](#)). In the Future of Jobs Reports from the World Economic Forum, which projects future skills needed from LinkedIn and Coursera data, at least 4 of the top 10 projected skills employers look for in 2025 will be non-cognitive ([Whiting, 2020](#)). These skills include learning strategies, initiative, leadership, and resilience.

Non-cognitive skill affects cognitive skill formation and many other important outcomes. Cognitive and non-cognitive skills are complements in the production of later ability formation ([Cunha & Heckman, 2007](#); [Cunha, Heckman, & Schennach, 2010](#)). Furthermore, non-cognitive skills have been shown to affect labor market ([Borghans, Meijers, & Ter Weel, 2008](#); [Heckman, Stixrud, &](#)

Urzua, 2006), health (Chiteji, 2010), criminal (Baker, Gruber, & Milligan, 2015), and macroeconomic outcomes (Balart, Oosterveen, & Webbink, 2018). Given how non-cognitive ability affects critical aspects of individuals' and nations' livelihoods, understanding the factors that lead to non-cognitive skill formation is a crucial undertaking.

This paper estimates the impact of peers on non-cognitive ability formation, specifically communication, leadership, and attitude. The setting for this study is the National Outdoor Leadership School (NOLS), wherein applicants are assigned to one of their top ranked preferences for course type. Importantly, this precludes students from self-selecting their peers. After being assigned to a course type and section, students traverse the wilderness with their group, narrowing down their peers to just their classmates.

Most randomized peers studies come from traditional college/schooling or armed forces institutions, but this setting is new and advantageous. Many other settings have relatively large groups, such as dormitory floors (Sacerdote, 2001), squadrons (Carrell, Hoekstra, & West, 2011b; Jones & Kofoed, 2020), or neighborhoods (List, Momeni, & Zenou, 2020), which means the actual peer reference group may only be a subset or also include those outside of the defined group.¹ This study is not vulnerable to this criticism, since these are relatively short time frames (mean = 37.99 days) with relatively small group sizes (mean = 10.54) and the courses are in the wilderness where the students have no contact with individuals other than classmates and instructor(s).

First, after conditioning on course indicators, there is no relationship between an individual's pre-course non-cognitive abilities and peer pre-course abilities. There is also no relationship along basic demographic characteristics. Given the absence of correlations among pre-treatment peer groups and individuals and considering students cannot select exact course, there is strong evidence that students are conditionally randomly assigned to peer groups.

Second, peers with higher non-cognitive ability causes students to rate themselves lower at the end of the course, conditional on the student's pre-course self-rating. This suggests that better peer groups leads to lower non-cognitive ability formation. Specifically, if the pre-course peer average

¹For example, this point about sorting within the defined peer group is made in Carrell, Sacerdote, and West (2013).

composite communication or leadership increases by 1, then the individuals rank themselves lower by about one-third of a communication or leadership question which is about 0.1 standard deviations. There are no effects of peer attitude on an individual's attitude. Allowing peer effects to be specified non-linearly results in the the magnitude of the estimates increasing (i.e. making them more negative).

Finally, the negative effects of peers depends on course and demographic characteristics. There are no subgroups where stronger peers lead to higher ability formation. When courses have more members, there are larger negative effects of peers on communication and leadership. For communication, there are negative effects on men and women and for above and below median age (19.58).

When courses are longer than median duration (30 days), peers have negative effects on communication; however, when courses are shorter than median duration, peers have negative effects on leadership. Peers have negative effects on leadership for those with prior NOLS experience and negative effects on communication for those without prior NOLS experience. Peers have negative effects on leadership for those who are above the median age and negative effects on the leadership abilities of females, but not males. Finally, there are no effects of peers on attitude for any subgroup.

Beginning with [Heckman and Rubinstein \(2001\)](#), the economics literature has paid increasing attention to the determinants and effects of non-cognitive ability. One important policy reason for being concerned with non-cognitive skill formation is that non-cognitive skills are believed to be malleable later in life, unlike IQ which is relatively constant by around age 10 ([Cunha, Heckman, Lochner, & Masterov, 2006](#), pg. 702).² Due to the greater malleability of non-cognitive skills later in life, investing in non-cognitive ability may have higher returns at later ages than investing in cognitive ability. Higher returns means investing in non-cognitive ability may pass cost-benefit analyses at ages where investing in cognitive ability would not.³ Another policy reason to be concerned with social skills is that they are increasingly valued in the labor market ([D. J. Deming,](#)

²This is due to the developmental cycle of the pre-frontal cortex ([Cunha et al., 2006](#), pg. 702).

³This is colloquially referred to as the Heckman Curve, see for instance [Rea and Burton \(2020\)](#).

2017; Weinberger, 2014) and the demand for them is expected to remain strong through at least 2025 (Whiting, 2020).

The determinants of non-cognitive ability is a relatively sparse area of research. This is perhaps due to the difficulty in finding measures of non-cognitive ability, whereas test scores are more readily available as a proxy for cognitive ability. Prior work has found that non-cognitive and cognitive ability are positively related to later non-cognitive ability (Cunha et al., 2010) and that parental inputs are a significant factor in the behavior of school-aged boys (Bertrand & Pan, 2013).

This paper adds to the literature on determinants of non-cognitive ability, by investigating peer effects as a potential cause of non-cognitive ability formation. Peer effects are especially difficult to estimate, because of endogenous peer group selection and the reflection problem (Manski, 1993). In this paper, endogenous peer selection is solved via random assignment to sections and the reflection problem is solved by using pre-treatment levels of non-cognitive ability.

The closest study to this one used students assigned to various classrooms in early education years, but found no effects of peers on non-cognitive abilities (Neidell & Waldfogel, 2010). Instead, we find that more able peers reduces non-cognitive ability formation. Our results may differ for several reasons, including differently aged subjects, different setting which allows for a precise definition of who the peer group is, and we have more empirical support that peers are randomly assigned. Also, the noncognitive abilities studied here are more widely varied.

While there is a large literature on how peers affect proxies for cognitive ability such as GPA and test scores (Sacerdote, 2011), there are also other literatures on how peers affect policy preferences (Boisjoly, Duncan, Kremer, Levy, & Eccles, 2006) and sociability of those with autism (Carter, Sisco, Chung, & Stanton-Chapman, 2010; Watkins et al., 2015). This study builds on the policy preferences literature by generalizing preferences over affirmative action to preferences for group objectives over one's own objectives. We find no evidence that this peer effect generalizes to a more general policy.

We build on how peers affect the sociability/communication of those with autism. Of the 85 studies synthesized in (Carter et al., 2010), there are only approximately 380 subjects or about 4

per study. Our sample size is over 3,700 students allowing for causal identification of peer effects. Furthermore, we extend this research question beyond just individuals with autism.

First, we discuss the setting of this study, the National Outdoor Leadership School. Next, we describe the data, measurement of non-cognitive ability and assignment of students to course section. Finally, we present the method, results, and end with a brief conclusion.

3.1 The National Outdoor Leadership School

We analyze data from a sample of students at the National Outdoor Leadership School (NOLS). NOLS was founded in 1965 to train skilled outdoor leaders, but soon broadened its focus to include non-professionals and novices. Since that time, it has grown into the largest backcountry permit holder in the U.S. with over 100,000 graduates.⁴

Courses and Skills Courses range from 14 days to a full year in length, though the most common are 30-day and semester-length courses. The courses focus on one or more of a diverse array of outdoor skills, including backpacking, mountaineering, river canoeing/kayaking, sea kayaking, rock climbing, and horsepacking.⁵ In addition to outdoor skills that vary by course, all NOLS courses target instruction to a common set of noncognitive skills including interpersonal and group communication, leadership, teamwork, and environmental ethics.

Students Most NOLS students are between the ages of 16 and 22 and most use the course for college credit. All NOLS courses are available for college credit that can be transferred to almost any college or university in the United States. The typical transfer student is awarded 16 semester credit hours for a NOLS Semester course.⁶ In addition to college-aged students, NOLS enrolls students from age 14 (where they are eligible for courses that are only offered to 14–15-year-olds)

⁴<https://www.nols.edu/en/about/history/>.

⁵Courses focusing on multiple skill areas are broken into sections in which one skill is emphasized. Most courses that are 30 days or fewer in length have one section and longer courses may have several sections.

⁶See <https://www.nols.edu/en/about/resources/college-credit/>.

to well above 40 (some courses are restricted to 40+).⁷ Also, NOLS goes to great lengths to ensure diversity in student enrollment. Outreach programs on college campuses and an extensive alumni representative network are used to reach potential students who may know about the school.

Cost NOLS awards more than \$800,000 in scholarships each year and attracts students from a wide range of educational and socioeconomic backgrounds.⁸ All courses are eligible for Federal Student Aid, AmeriCorps benefits, Veterans Affairs benefits, and in some cases employer tuition benefits. The total cost (without financial aid), including tuition, travel, room, board, college transfer credit expenses, gear, and personal expenses, of a popular semester-length course, the Semester in the Rockies, is about \$12,400 in 2008. This cost compares favorably to the costs of a semester at a public four-year college or university and is considerably lower than the cost of a private university (Baum & Ma, 2007).

3.2 Data

The administrative data comes from NOLS and the non-cognitive ability measures are from a survey developed by a research team affiliated with NOLS and the University of Utah called the Student Outcome Assessment Project (SOAP). The survey was designed to assess the instructional effectiveness of various course characteristics. It was administered to about 4,000 students in various courses over a two year period in 2005 and 2006. The sample in this paper includes 3,773 students, 87 course types, and 358 course sections.

Individual Level At the individual level, demographics, prior experience, and self-rated ability are observed. The demographic characteristics are age (in days), race (6 categories), and gender. Prior experience is a binary variable that equals 1 if a student has taken a course at NOLS in the

⁷NOLS also conducts custom and professional training courses for schools, organizations, and corporations including NASA, the Kellogg School of Business, the Wharton School, the United States Naval Academy, Google, and Fidelity Bank and Trust, groups arguably very concerned with human capital formation and productivity, which strongly suggests that the skills taught at NOLS are important determinants of educational and labor market success.

⁸See <https://www.nols.edu/en/expeditions/planning/financial-aid/>.

past. The individual-level data also contains instructor assigned grades for the course (from F-A+), but this information is missing for 682 students (18.1%) in the sample.⁹ Finally, the data contains which course and which section of that course the student participated in.¹⁰

Course Level and Section Level At the course type level, the course location and other course details are observed. Course details include whether a course required technical skills and another is whether a course has cultural elements. At the course-section level, the dates of beginning and ending and the instructor(s) are observed.¹¹ On average, there are 4.11 sections per course type.

3.2.1 Non-Cognitive Skills

The outcomes of interest are the noncognitive skills of communication, leadership, and attitude.¹² The actual questions, measuring various aspects of these abilities and chosen to measure targeted outcomes in the NOLS curriculum, are shown in Table 3.1.¹³ To each question, students rated themselves on a scale of 1 (worst) to 8 (best). Section 3.3.2 discusses how potential measurement error, from subjective self-ratings, affects the results.

The survey underwent several pilot trials prior to being implemented to ensure that the survey questions adequately and consistently measure the associated abilities.¹⁴ Surveys were administered within two hours of students returning from the field, typically at the end of the course. For semester-length courses, surveys were conducted during a convenient break between parts close to the 30-day time frame of the typical NOLS course.

At the same time these responses were collected, students also *retrospectively* reported their skills as they were just before the course began. This retrospective structure was chosen to avoid

⁹The data goes D-, D, D+, C-, C, C+, B-, B, B+, A-, A, A+. The grade data is recoded so that “A+” is a 12 and “D-” is 1. While there are “grade points” associated with each letter, the recoding is preferred so that differences in grade point differences between grades do not drive results (i.e. B (3) to B+ (3.3) is different than B+ (3.3) to A- (3.7)).

¹⁰Unfortunately, the data on how students ranked their preferences over courses is incomplete and cannot be used to investigate how many students received their top preference, second preference, etc.

¹¹At the instructor-course section level, the age and experience of the instructor(s) are observed. Instructor data is not used.

¹²The student survey is reproduced in its entirety in Figures C.1 and C.2.

¹³NOLS’s definition of these skills are available at: <https://blog.nols.edu/topic/leadership>.

¹⁴See Sibthorp, Paisley, and Gookin (2007) for details.

response-shift bias, a serious problem observed in pilot trials and in many surveys of pre-treatment levels of instruments. This bias occurs when the treatment itself changes the respondent's metric of evaluating the instrument (Howard, 1980; Sibthorp et al., 2007).¹⁵ This retrospective structure of the pre-treatment measures raises the concern that students pre-treatment scores are endogenous to the peer group; however, results presented in Section 3.3.2 find no evidence that this is the case. In fact, this pre-treatment measure helps address the reflection problem in peer effect estimations (Manski, 1993), by measuring ability before the group interacted.

Combining Survey Questions Into Single Measures of Ability There are 4 questions that serve as the basis for each non-cognitive ability, but the outcomes and treatments of interest are the single underlying abilities. To simplify the analysis, the answers for each of the 4 survey questions, for each skill, are combined into a single measure of that skill. To combine the questions, an average is taken across each of the 4 questions, for each skill, for each individual.¹⁶ Robustness checks regress each individual post-course ability question on pre-course peer ability and find results are not sensitive to this aggregation method. Throughout the paper, each average of 4 questions is referred to as composite communication, leadership, and attitude respectively.

3.2.2 Descriptive Statistics

Table 3.2 presents descriptive statistics of student demographics, student pre and post cognitive ability, course type attributes, and course section compositions. Panel A shows the sample is composed of 64% males, aged 21.14 years old on average, and 52% have taken a prior course at NOLS. A majority, 91%, are white; however, race information is missing for over 300 individuals which leads to them being dropped in regressions that include race.

¹⁵Also, the uniform time frame of surveying students approximately 30 days into the course mitigates a problem observed in pilot trials where pre-treatment scores would creep up as time passed. Similar to “teacher’s bias” students begin to forget what it was like to not know the things that they now know very well.

¹⁶An alternative method for interpreting the information is principal components analysis (PCA), which has the unfortunate drawback of being difficult to interpret in a regression equation. Nevertheless, subsection C.2 performs PCA and finds that each question loads on the same component for individual skills and after the leave-out means are calculated within groups.

Panel B presents individual's composite communication, leadership, and attitude from before the course.¹⁷ On average, students rate themselves around a 5.5 for each composite ability. While the questions are on an integer scale, the process of averaging 4 questions leads to decimal values. The standard deviation is one-fifth of the mean, suggesting adequate variation in pre-course ability for regression analysis. The means are also close to the medians, suggesting the distributions are approximately normally distributed.

Panel C presents the same information for after the course and the instructor assigned grades. On average, the students rate themselves around 6.5 for each composite ability, an improvement from the pre-course average of 5.5. The standard deviation is about one-third smaller than the standard deviation in the the pre-course answers, but the distributions are still approximately normally distributed. Students receive a B+ on average in the course and the standard deviation is about 1.5 grades (B- to A).

Panel D shows course type characteristics. There are some course types with missing data, so the following interpretations are for courses that have non-missing data for that characteristic. Most of the courses are in the summer and in the United States. A quarter of the course types are technical and 7 percent have a cultural component. On average courses last for 37.99 days, but that is slightly skewed upwards with the median length being 30 days. Panel E shows course section level composition. On average, the proportion of students in courses that are males is 64% and that are white is 82%. The average number of students in a section is 10.54, with a median of 11.

Peer Assignment and Course Selection Students do not self-select their peers, though they do select which course to take. They rank three courses in order of preference, and they are not always assigned their first choice. Students have no control over which group they are assigned to.¹⁸ For

¹⁷Descriptive statistics on the disaggregated questions are available in [Table C.1](#). Similar conclusions can be drawn.

¹⁸According to NOLS admissions, there are rare exceptions, such as friends taking a course together who wish to be assigned to the same group, or siblings taking a course at the same time who may wish to be assigned to different groups. A NOLS admissions officer stated that “very few, less than five percent” of students know their potential peers prior to enrolling in the course. Another situation where assignment may be non-random is when NOLS admissions makes an effort to ensure that any one female (or in very rare cases, male) is not the only female in a group. This means that if a course has only one female enrolled, the admissions officers will either attempt to find another or offer her the option to switch to a different course. Thus the characteristics of lone females in a group may differ

most courses the gender ratio is determined by enrollment, but for popular courses with multiple sections, the admissions staff forms some groups that are roughly half male and the remaining males are sorted into all-male groups (note the bump on the right tail of [Figure C.3](#)). Furthermore, the non-cognitive abilities are not made available to NOLS at the time of assignment to course sections, because the survey is administered at the end of the course.

3.3 Method

To estimate effects of pre-course peer ability, OLS, linear-in-means models are used. The models are specified,

$$Ability_{i,t+1} = c + \beta \overline{Ability}_{-i,t} + \gamma Ability_{i,t} + X_i \psi + \mu_C + \phi_Y + \theta_M + e_i. \quad (3.1)$$

The i stands for individual, $-i$ is for everyone in the course section besides person i , t is prior to the course, $t+1$ is after the course, C is for course type, Y is for course section year, and M is for course section month. Standard errors are clustered by course section to account for correlated errors within course sections and potential model mis-specification.

3.3.1 Peer Effect Specification

The coefficient of interest, representing the effect of peers, is β . It represents the estimate for the leave-out average pre-course peer effect on individual i 's post course ability. The leave-out average peer effect is calculated as the average ability of all of individual i 's coursemates, excluding individual i . The leave out pre-course composite peer ability is calculated,

$$\overline{Ability}_{-i,t} = \frac{\sum_{j \neq i}^{n_g} A_{j,t}}{n_g - 1},$$

systematically from the rest of the sample because they have chosen not to select out of the group. Females in a group that is otherwise all-male are included in the sample, but results are nearly identical if I drop these students from the sample.

in which subscript g stands for group and A for ability.

While Equation 3.1 specifies peer ability as linear-in-means, this may lead to mis-specification error. The potential mis-specification is due to many students near the upper bound of the composite abilities (8), as shown by Figure C.4. Furthermore, there may be non-linear effects of peer ability. Robustness checks will address potential mis-specification and non-linear effects using non-parametric kernel regression.

Individual Pre-Course Ability Controls An important control variable is individual i 's pre-course ability. Students with different initial ability may be differently affected by their peers. By including this control, estimates compare those of the same pre-course ability that are assigned to higher or lower skilled peer groups.

Demographic Controls Other individual level pre-treatment covariates are included in X . The demographic variables included are continuous age (in days), 6 indicator variables for different ethnicity categories, and a binary variable for male. The final variable is a binary variable for whether or not the individual has previous experience in a NOLS course.

Course Section Timing Fixed Effects The last fixed effects are for course month, θ_M , and course year, ϕ_Y . These indicators control for trends in the types of individuals who select into NOLS over time. Also, these control for the potential that weather may be correlated with peer effects and/or ability formation.

3.3.2 Identification

Finally, there are fixed effects that account for differences in students who select different course types. The μ_C is a vector of dummy variables for course types. These indicators ensure that estimated peer effects are not confounded by individuals selecting into different course types. For instance, those with stronger leadership skills may be more likely to take a course requiring technical skill; however, these course indicator variables allows β to be estimated across individuals

who are in the same course type, but different course section with different peer group ability. This is the same strategy used in [Feld and Zölitz \(2017\)](#) to control for students in the same course type, but different sections.

One might be concerned about the connections that [J. D. Angrist \(2014\)](#) makes between peer effects estimations and weak instrumental variable (IV) designs. One of the main conclusions of [J. D. Angrist \(2014\)](#) is that β is equivalent to a 2SLS model, in which group indicators are instruments for peer group ability, and an OLS model. These coefficients may differ due to peer effects or any other reason. One prominent concern is measurement error which attenuates IV estimates less than OLS estimates ([Feld & Zölitz, 2017](#), pg. 391). This is not a concern, because if students are randomly assigned to sections and one controls for course, then it is appropriate to consider peers randomly assigned and any potential measurement error leads only to attenuation bias ([Feld & Zölitz, 2017](#), pg. 393).

Conditional Randomization

[Table 3.3](#) investigates whether controlling for course type makes the assignment of students to peer groups as good as random. The analysis regresses individual pre-course abilities and demographics on pre-course group ability and demographics. If there is no relationship between pre-course group characteristics and demographics and individuals' own skills and demographics, then it is consistent with peers being as good as randomly assigned.

First, Panel A begins with simple bivariate relationships. As can be seen, when not controlling for course type, all group demographics and skills are correlated in a statistically significant way with individual characteristics. For instance, if leave-out group pre-course composite communication increases by 1, then individual pre-course composite communication increases by 0.41.

When course type indicators, which control for different individuals selecting into different course types, are included, all the relationships get closer to 0 and 5 out of 6 become statistically insignificant. For instance, after controlling for course type, if the group becomes more skilled at communication, leadership, attitude or is more male or more experienced, these have no statisti-

cally significant relationship to an individual's own skill, gender, or experience.^{19,20}

The one coefficient that remains statistically significant is group age. This is likely due to age restrictions on courses that are decided by NOLS which leads to discontinuities in the ages that are eligible, making global linear regression mis-specified.²¹ Table C.2 shows that while age does have a global linear effect, there is no relationship of peer age on individual age within the age groups of 14-15 and 16-22, compared to 22 and above. Overall, it is reasonable to conclude that peers are randomly assigned, after conditioning on course types, so that β can be considered a causal effect of peers. This conditional randomization addresses the selection problem which is associated with estimating peer effects (Manski, 1993) and makes it reasonable that potential measurement error in peer ability only causes attenuation bias, after conditioning on course type (Feld & Zölitz, 2017).

3.4 Results

This section begins by showing binned scatter plots. Afterwards, OLS models and non-parametric kernel regression are used. Next, heterogeneous effects are estimated by gender, previous experience, number of days, and number of members.

3.4.1 Binned Scatter Plots of Peer Ability on Individual Ability

Next, the binned scatter plots investigating the effects of pre-course peer ability on post-course individual ability are shown in Figure 3.1. All Figures control for own pre-course ability, so that comparisons are made between those with equal pre-course own ability, but different pre-course peer ability. According to methods in Cattaneo, Crump, Farrell, and Feng (2019), all Figures also use the number of bins that are optimal from a bias-variance tradeoff.

In the first row, Figures 3.1a and 3.1b show that there is a negative correlation between pre-

¹⁹Sex is uncorrelated partly because of the offsetting effects of efforts to balance the gender ratios of course sections.

²⁰The way conditioning on course types removes the relationship between peer pre-course ability and own pre-course ability can be seen in the binned scatter plots, one set bivariate and one set conditional on course type indicators, in Figure C.5.

²¹Figures C.6 and C.7 show these discontinuities around age 16 and 22.

course peer composite communication and after course own composite communication. In the second row, Figures 3.1c and 3.1d show the same negative correlation for composite leadership. Figures 3.1e and 3.1f are similar for attitude, although the linear fit is not as close to all of the data points.

Figures 3.1a, 3.1c, and 3.1e find negative associations, holding own pre-course ability constant. Additionally, Figures 3.1b, 3.1d, and 3.1f all control for course type. Since Table 3.3 suggests peers are as good as randomly assigned when conditioning on course type, these Figures support a causal interpretation of this negative association.

3.4.2 OLS Regressions of Peer Effects on Noncognitive Abilities

The results from estimating Equation 3.1, the effect of peer pre-course composite ability on individual post course composite ability, are presented in Table C.6. Columns 2 and 3 use less observations, due to missing race data for some students. Overall, the independent variables explain a moderate amount of variation in the dependent variables with R-squared between 0.420 (Panel C, column 1) and 0.521 (Panel A, column 3).

Panel A estimates how pre-course peer average composite communication affects post-course individual composite communication. In Panel A, column 1, a 1 level increase (i.e. from 3-4, 4-5, etc.) in pre-course peer average composite communication reduces post-course individual communication by 0.14, for individuals with the same pre-course communication ability. This effect is statistically significant above 99% confidence. Panel A, column 2 adds individual covariates and the effect gets slightly smaller, but remains statistically significant.

Two concerns with estimates in Panel A, columns 1 and 2 are potential measurement error in pre-course peer average composite communication and non-random peers (Feld & Zölitz, 2017). Table 3.3 in Section 3.3.2 suggests that after conditioning on course type, controlling for different individuals selecting into different courses, that peers are as good as randomly assigned. When peers are as good as randomly assigned, measurement error only results in attenuation bias (Feld & Zölitz, 2017). So, column 3 includes indicators for course type and the effect size becomes

smaller in magnitude, but remains statistically significant. The effect is reasonably sized, it is interpreted that a 1 question increase in pre-course peer average composite communication reduces an individual's post course communication by about one-third of one question, which is about one-tenth of a standard deviation in post-course individual composite communication.²²

Panel B investigates how pre-course peer average composite leadership affects post-course individual composite leadership. Columns 1 and 2 find that higher pre-course peer average composite leadership reduces post-course individual composite leadership by 0.13 and 0.11, for those of the same pre-course composite leadership, and both are statistically significant above 99% confidence. Panel B, column 3 includes course type indicators and finds that the estimated effect decreases to 0.07, but remains statistically significant above 95% confidence. This estimate is nearly equivalent to the estimate in Panel B, column 3, implying a lower post-course individual composite rating by about one-third of one question which is also about one-tenth of a standard deviation.

Panel C investigates how pre-course peer average composite attitude affects post-course individual composite attitude. In column 1, an increase in peer attitude reduces one's own individual post-course composite attitude, for individuals with similar pre-course composite attitude. However, once demographic controls (column 2) and course type indicators (column 3) are added, the estimated effects are no long statistically distinguishable from 0 at conventional levels.²³

The consistently estimated negative effects are somewhat puzzling on the surface. Typically, better peers have increased individual grades in the literature (Sacerdote, 2011). Table C.3, which estimates each how peer's answer to each individual question affects each individual post-course question, suggests it is not due to the combination of the 4 questions into 1 composite measure, because all the estimates are of similar magnitudes with no obvious outliers. Two potential hypotheses for the negative results are specification error from linear-in-means models with outcomes that

²²Because it is an average of 4 questions, a reduction of 1 would imply they scored 1 lower on each question or 4 lower on question.

²³One might believe that reducing measurement error in the dependent variable by using a non-self reported outcome would lead to different results. Instructor-assigned grades are observed, so Figure C.9 and Table C.6 investigate peer non-cognitive ability on grades. Figure C.9 seems to suggest that there is some moderate, positive effect of peer non-cognitive ability on grades; however, this becomes much less precise once course type is controlled for. Table C.6 finds no statistically significant effects of peer non-cognitive ability on grades.

are near the boundary of the support and heterogeneous effects. The first step to investigate this is to use non-linear models of peer effects and the second step is to investigate heterogeneity.

3.4.3 Non-Linear Peer Effects

In [Equation 3.1](#), it is assumed that for each 1 unit increase in pre-course peer average composite ability, there is a constant effect on post-course individual composite ability. However, it is simple to envision that going from a group with 2 to a group with a 3 may have differently sized effects than going from a group with a 7 to a group with an 8.²⁴ In order to better characterize non-linearities, the linear effect of peers must be relaxed.

To relax linearity, non-parametric kernel regressions are estimated. The equations of these models follow,

$$Ability_{i,t+1} = f(\overline{Ability}_{-i,t}, Ability_{i,t}, X_i, \mu_C, \phi_Y, \theta_M) + e_i, \quad (3.2)$$

in which all subscripts and variables are the same as [Equation 3.1](#). The method does not assume that $f(\dots)$ is linear or linear in parameters and is estimated using local linear regression. Continuous covariates use an Epanechnikov kernel and discrete covariates use the Li-Racine Kernel ([Li, Lu, & Ullah, 2003](#)). As recommended by [Cattaneo and Jansson \(2018\)](#), standard errors are calculated by bootstrapping.

Average Effects From Regression The average effects, which are averages of the derivatives of the estimated conditional mean function, are presented in [Table C.4](#). Across all abilities, the average effect of peers is more negative than estimated by linear regression. It is likely that this is due to non-linear effects of peer ability, causing the linear specification to underestimate the true average effect.

²⁴The various justifications for non-linear, heterogeneous effects are detailed in [Table .](#) in [Sacerdote \(2011\)](#).

3.4.4 Marginal Effects

Figure 3.2 shows the estimated conditional mean functions from Equation 3.2. The left sides are trimmed off, because there are few observations close to 0. The confidence intervals are not shown, because they make the scale on the graph hard to see.

In Figure 3.2a, there's a downwards slope from 5 to 6 before it levels off. The increase from 4 to 5 is due to few observations, so it can be safely ignored. In Figure 3.2b, there is a constant downwards slope over the whole support of peer leadership. In Figure 3.2c, there is no discernible slope.

3.4.5 Heterogeneous Effects

It is important to understand the channels peer ability negatively affects ability formation. Furthermore, are there some subgroups for which peers exert a positive influence? First, this section estimates peer effects separately by course characteristics; second, effects are estimated separately for different demographic characteristics and experience with NOLS. All models are estimated using OLS, which Section 3.4.3 suggests will lead to negative effects that are closer to 0 due to linearity being a mis-specification, so these can be viewed as conservative.

Heterogeneous Results by Course Type Characteristics

First, effects are estimated separately by course and section characteristics, in particular by the number of members within a course section and the number of days a course lasts. The specifications use medians to split the data into long versus short courses and many versus fewer members. Using the median results in an even split which results in larger sample sizes above and below the median. Table 3.5 shows the heterogeneous results by course characteristics. There are no statistically significant positive effects anywhere, but there is heterogeneity by both number of section members and course length.

Course Members First, columns 1 and 2 of [Table 3.5](#) show effects for subsamples by whether or not there are above median (11) section members. Panel A and B shows that for communication and leadership, being in a larger course section and better peers reduces individuals' ability by 0.15 and 0.12 respectively which is a lower rating on one half question. These are larger than the effects shown in [Table C.6](#). There are no statistically significant effects of peers on ability formation for attitude (Panel C) or for courses with less than median members (column 2, Panel A, B, and C). The larger groups having a larger effect is consistent with the mechanism being comparing yourself to more able coursemates.

Course Length Second, columns 3 and 4 of [Table 3.5](#) show effects for subsamples by whether or not the courses are longer than median duration (30 days). For communication (Panel A), peers have larger and statistically significant (above 99% confidence) effects in shorter courses (column 4), but no effect in longer courses (column 3). For leadership (Panel B), peers have larger negative effects (0.09) in longer courses (column 3) that are statistically significant at 90% confidence. The effect of peers in shorter courses is 0.05 (column 4), but p-value is 0.124, so it is not statistically significant at conventional levels. There are no effects of peers on attitude for short or long courses (Panel C, columns 3 and 4). It is interesting that peers have negative effects on communication in short courses and negative effects on leadership in longer courses.

Heterogeneous Results by Demographics

Next, the effect of peers are estimated separately by demographic characteristics. The demographic characteristics are gender, age, and previous experience.²⁵ [Table 3.6](#) presents the separate estimates of peer effects by demographic characteristics. Again, there are no positive, statistically significant effects of peers for any subgroup.

²⁵There are too few non-white students to have meaningful estimates of peer effects for white and non-white individuals.

Gender Columns 1 and 2 inspect whether there are different effects of peers on men or women. For communication, there are negative and statistically significant effects of peers on men and women and the effect is slightly larger for women. For leadership, there is no negative effect of peers on men, but there is a large negative effect on women which is statistically significant at 99% confidence. For attitude, there are no effects of peers on men or women.

Age Columns 3 and 4 investigate whether there are different effects of peers for older versus younger individuals. There is an approximately equal negative effects of peers on communication for both older (0.09, p-value = .028) and younger (0.07, p-value = .090) individuals. For leadership, there is a large negative effect (0.10, p-value = 0.003) of peers for individuals above the median age, but no effect on those who are younger. It is somewhat surprising that those who are older are more influenced by their peers, but not unreasonable given that non-cognitive skills remain malleable for longer (Cunha et al., 2006, pg. 702). Finally, there are different effects of peers on attitude by age.

Previous Experience Columns 5 and 6 investigate whether there are different effects of peers by whether or not the individual had previous experience in a NOLS course. It might be these case that those who have experience are more able to learn from their peers or are already experienced and conflict with newer group members. For communication, those who have prior experience are negatively affected by their peers, reducing their score by about half of a question (statistically significant at 99% confidence.) For leadership, those who have prior experience are not harmed by better peers, but those without prior experience have 0.1 lower post-course composite leadership scores (statistically significant at 95% confidence). There are, again, no effects of peers on attitude. It is interesting that those with prior experience have lower communication scores due to better peers, but those with no prior experience have lower leadership scores due to better peers.

3.5 Conclusion

Non-cognitive abilities are increasingly rewarded in the labor market and have been shown to be an important factor in many critical aspects of human life. There is a sparse literature examining the determinants of non-cognitive ability, which finds that parents and prior ability lead to higher non-cognitive ability. One paper has examined the effect of peers on non-cognitive ability ([Neidell & Waldfogel, 2010](#)), but it finds no effects.

This paper investigates the causal effect of peers on noncognitive ability in a setting, the National Outdoor Leadership School, that addresses some of the major impediments to estimating causal effects of peers and measurements of non-cognitive ability. Peers with better communication or leadership skills reduces the same skill by about one-tenth of a standard deviation. These effects differ in their sizes for subgroups, including course length, course members, gender, age, and previous experience.

There are a couple limitations. First, non-cognitive ability is self-reported. Other studies use behavioral incidents as a more objective measure.²⁶ Second, there is no information on socioeconomic status such as income, which would be interesting to have given the importance of parental inputs mattering for disruptive male behavior in early education ([Bertrand & Pan, 2013](#)).

These results have significant bearing on policy-making. Many studies have found that increasing peer quality can have positive effects on proxies for cognitive abilities such as GPA or test scores, these results suggest that this increase in cognitive ability comes at the detriment to critical non-cognitive abilities: communication and leadership. Furthermore, stronger peers reduce women's leadership, suggesting that peers could play a role in the wide gap in gender representation in start up funding and executive boards.²⁷ That being said, it may be difficult to address the negative effect of peers on women given findings that suggest peer groups form endogenously, even in controlled environments ([Carrell et al., 2013](#)).

²⁶Still, behavioral incidents are a coarse measure of non-cognitive ability.

²⁷In 2019, less than 3% of all venture capital investment went to women-led companies and only 20% of U.S. venture capital went to startups with at least woman on the founder team ([Aleman, 2020](#)). In 2018, women held 16.9% of executive board positions worldwide ([Deloitte, 2018](#)).

There are several interesting directions for future research. First, the sample comes from a select population, those who participate in the National Outdoor Leadership School. This population may be fundamentally different than other populations, so extending these results to other populations would be invaluable. Second, are there interactions amongst formation of certain non-cognitive abilities? Finally, do age rank or demographic-specific peer effects matter?

Table 3.1: Composition of Noncognitive Ability Measures

Students are asked to rate how well the following statements describe them. Composite ratings for each skill are computed as the mean of the responses in each category.

COMMUNICATION

1. I can lead others in a discussion.
2. I give constructive feedback.
3. I am good at reading other people's body language.
4. I express my ideas clearly

LEADERSHIP

1. I take initiative in completing group tasks
2. I often take responsibility without being asked
3. I am good at making decisions
4. I make decisions in a timely manner

ATTITUDE

1. I am patient with others
 2. I place emphasis on group goals above personal goals
 3. I maintain a positive attitude in adverse conditions
 4. I can manage conflict that occurs between group members
-

Note: These questions are chosen to measure targeted outcomes in the NOLS curriculum. To each question, students rated themselves on a scale of 1 (worst) to 8 (best). The survey underwent several pilot trials prior to being implemented to ensure that the survey questions adequately and consistently measure the associated abilities. Surveys were administered within two hours of students returning from the field, typically at the end of the course. For semester-length courses, surveys were conducted during a convenient break between parts close to the 30-day time frame of the typical NOLS course. At the same time these responses were collected, students also *retrospectively* reported their skills as they were just before the course began. This retrospective structure was chosen to avoid response-shift bias, a serious problem observed in pilot trials and in many surveys of pre-treatment levels of instruments.

Table 3.2: Descriptive Statistics

Panel A: Demographics						
	Mean	Median	SD	Min	Max	Count
Male	0.64	1.00	0.48	0.0	1.0	3773
Age	21.14	19.58	6.49	13.6	70.4	3706
White	0.91	1.00	0.29	0.0	1.0	3403
Previous Experience	0.52	1.00	0.50	0.0	1.0	3773
Panel B: Pre-Course						
	Mean	Median	SD	Min	Max	Count
Composite Communication	5.48	5.50	1.17	1.0	8.0	3709
Composite Leadership	5.57	5.75	1.28	1.0	8.0	3708
Composite Attitude	5.37	5.50	1.21	1.0	8.0	3702
Panel C: After Course						
	Mean	Median	SD	Min	Max	Count
Composite Communication	6.49	6.50	0.83	1.0	8.0	3693
Composite Leadership	6.68	6.75	0.84	1.0	8.0	3713
Composite Attitude	6.48	6.50	0.89	1.0	8.3	3698
Instructor Grade	9.05	9.00	1.53	2.0	12.0	3091
Panel D: Course Level						
	Mean	Median	SD	Min	Max	Count
Course in US	0.59	1.00	0.50	0.0	1.0	87
Technical Course	0.25	0.00	0.44	0.0	1.0	60
Cultural Course	0.07	0.00	0.25	0.0	1.0	60
Summer Course	0.80	1.00	0.40	0.0	1.0	60
Course Days	37.99	30.00	22.89	9.0	94.0	77
Panel E: Section Level						
	Mean	Median	SD	Min	Max	Count
Proportion Male	0.64	0.64	0.19	0.0	1.0	358
Proportion White	0.82	0.82	0.14	0.0	1.0	358
Number of Members	10.54	11.00	2.63	2.0	22.0	358

Note: Note: Data is on National Outdoor Leadership School (NOLS) students from 2005-2006. In Panel A, previous experience means that the student has participated in a NOLS course in the past. Composite abilities are averages of the 4 questions that make up each ability as displayed in [Table 3.1](#). Panel B is from before the course begins. Panel C is from the end of the course. Instructor grade is a conversion from letter grades to an integer scale from 2-12. Panel D shows course level characteristics. Panel E shows course section level composition information.

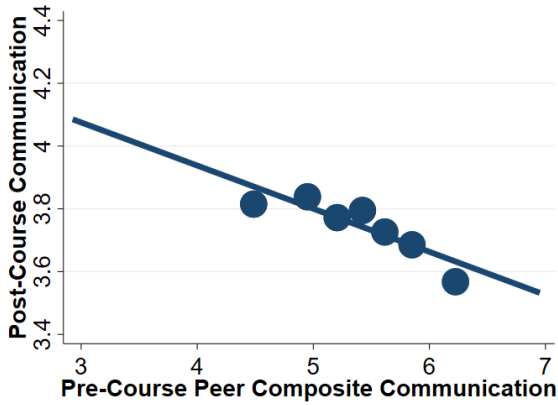
Table 3.3: Own Characteristics Regressed onto Peer Characteristics

Panel A: Bivariate						
	(1)	(2)	(3)	(4)	(5)	(6)
	Communication	Leadership	Attitude	Age	Male	PrevExp
Peer Leave-Out Average	0.41*** (0.03)	0.48*** (0.03)	0.40*** (0.04)	0.93*** (0.01)	0.33*** (0.05)	0.46*** (0.04)
Observations	3709	3708	3702	3706	3773	3773
Panel B: Course Characteristics						
	Communication	Leadership	Attitude	Age	Male	PrevExp
Peer Leave-Out Average	-0.02 (0.07)	-0.10 (0.07)	-0.02 (0.07)	0.40*** (0.10)	-0.06 (0.10)	0.00 (0.08)
Observations	3709	3708	3702	3706	3773	3773

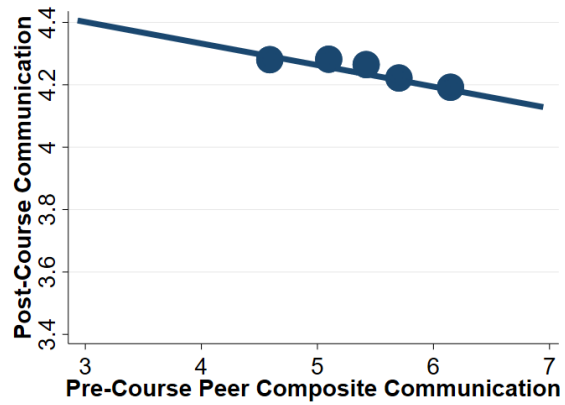
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by courseid, in parentheses. Panel A includes no controls. Panel B includes fixed effects for course type indicators, course section years, and course section months. The dependent variable in columns 1-3 are individual's pre-course composite communication, leadership, and attitude respectively. In column 4, the dependent variable is an individual's age. In column 5, the dependent variable is a binary variable for whether the student is male. In column 6, the dependent variable is a binary variable for whether the student had previous experience in a NOLS course. The independent variable is the corresponding leave-out averages from the individual sectionmates.

Figure 3.1: The Effect of Pre-Course Peer Average Composite Ability on Post-Course Own Composite Ability

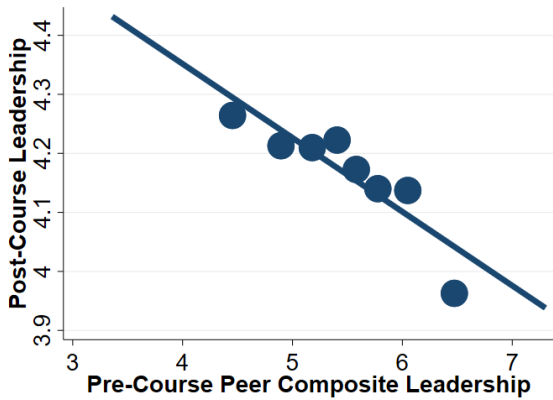
(a): Communication, Pre-Course Ability Control



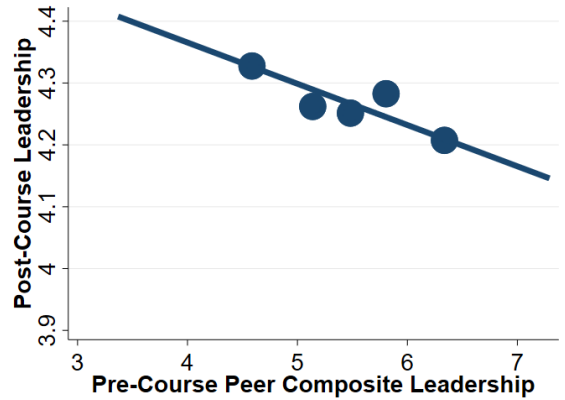
(b): Communication, With More Controls



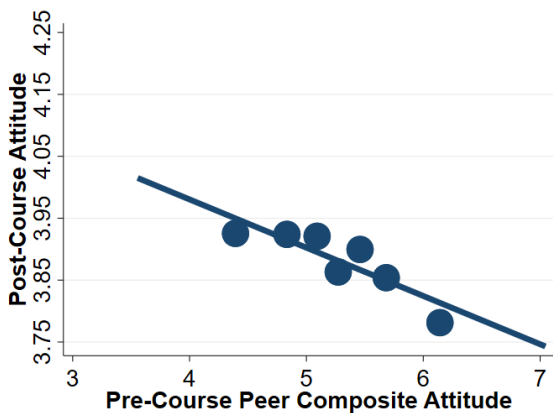
(c): Leadership, Pre-Course Ability Control



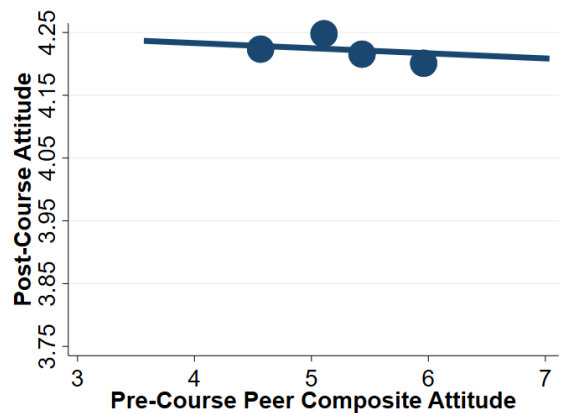
(d): Leadership, With More Controls



(e): Attitude, Pre-Course Ability Control



(f): Attitude, With More Controls



Note: Y-axis is after course individual composite ability. X-axis is pre-course peer average composite ability. Number of bins is optimal for variance-bias tradeoff (for integrated mean square error), as defined by Cattaneo, Crump, et al. (2019). Panel A, C, and E are conditional on peer group average and own ability before the course. Panels B, D, and F include course indicators, gender, race, and age as covariates. All figures control for individual pre-course ability. In A, N=3,661. In B, N=3,265. In C, N=3682. In D, N=3286. In E, N=3669. In F, N=3278.

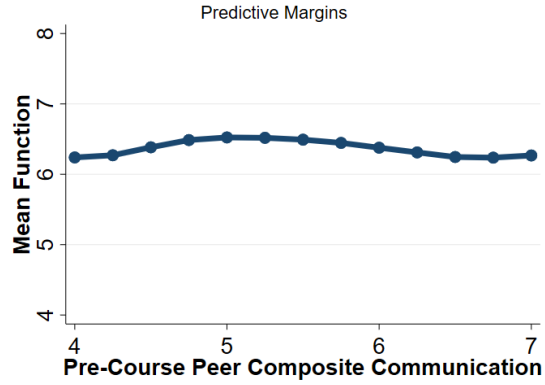
Table 3.4: OLS of Peer Group on Noncognitive After and Improvement

Panel A: Communication			
	(1)	(2)	(3)
Pre-Course Peer Average Composite Communication	-0.14*** (0.02)	-0.11*** (0.02)	-0.08*** (0.03)
Observations	3661	3265	3265
R-Sq	0.478	0.483	0.521
Individual Covariates		X	X
Course Indicators			X
Panel B: Leadership			
	(1)	(2)	(3)
Pre-Course Peer Average Composite Leadership	-0.13*** (0.02)	-0.11*** (0.02)	-0.07** (0.03)
Observations	3682	3286	3286
R-Sq	0.454	0.457	0.488
Individual Covariates		X	X
Course Indicators			X
Panel C: Attitude			
	(1)	(2)	(3)
Pre-Course Peer Average Composite Attitude	-0.08*** (0.03)	-0.05 (0.03)	-0.00 (0.03)
Observations	3669	3278	3278
R-Sq	0.420	0.422	0.454
Individual Covariates		X	X
Course Indicators			X

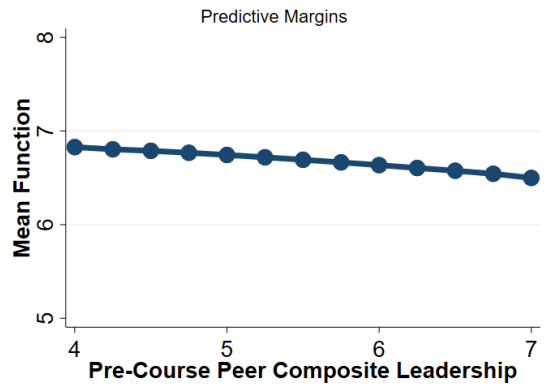
Note: Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by courseid, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual composite communication and the independent variable is pre-course peer average composite communication. In Panel B, the dependent variable is post-course individual composite leadership and the independent variable is pre-course peer average composite leadership. In Panel C, the dependent variable is post-course individual composite attitude and the independent variable is pre-course peer average composite attitude. All columns control for the individual's pre-course ability for the outcome of interest in that panel. Column 1 includes no additional covariates. Column 2 adds controls for male, age, and race. Column 3 adds indicators for course type and fixed effects for course section month and year.

Figure 3.2: Marginal Effects From Non-Parametric Kernel Regressions

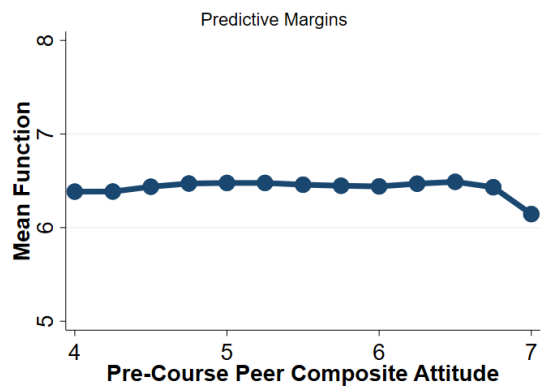
(a): Communication Margins



(b): Leadership Margins



(c): Attitude Margins



Note: Figures show the predictive margins of the mean function from the non-parametric kernel regression shown in Equation 3.2 and average effects presented in column 2 of Table C.4. Standard errors make it too difficult to interpret, so they are shown in Figure C.8. The standard errors are typically only a few hundredths wide.

Table 3.5: Heterogeneity in Peer Effects by Course Characteristics

Panel A: Communication				
	Course Members		Course Length	
	(1)	(2)	(3)	(4)
	> Median	≤ Median	> Median	≤ Median
Pre Comm. Leave-Out Mean	-0.15*** (0.05)	-0.05 (0.04)	0.02 (0.07)	-0.10*** (0.03)
Observations	1336	1929	1065	2058
R-Sq	0.511	0.549	0.479	0.544
Panel B: Leadership				
	Course Members		Course Length	
	(1)	(2)	(3)	(4)
	> Median	≤ Median	> Median	≤ Median
Pre Lead Leave-Out Mean	-0.12*** (0.04)	-0.03 (0.04)	-0.09* (0.05)	-0.05 (0.03)
Observations	1334	1952	1077	2065
R-Sq	0.516	0.484	0.468	0.487
Panel C: Attitude				
	Course Members		Course Length	
	(1)	(2)	(3)	(4)
	> Median	≤ Median	> Median	≤ Median
Pre Att. Leave-Out Mean	-0.05 (0.06)	0.01 (0.04)	-0.02 (0.05)	0.01 (0.04)
Observations	1339	1939	1066	2067
R-Sq	0.469	0.456	0.431	0.462

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual composite communication and the independent variable is pre-course peer average composite communication. In Panel B, the dependent variable is post-course individual composite leadership and the independent variable is pre-course peer average composite leadership. In Panel C, the dependent variable is post-course individual composite attitude and the independent variable is pre-course peer average composite attitude. All columns control for the individual's pre-course ability for the outcome of interest in that panel. Additional controls in all columns include age (in days), race indicators, and . The median number of course members is 11 and the median number of days is 30.

Table 3.6: Heterogeneity in Peer Effects by Demographics

Panel A: Communication						
	Gender		Age		Previous Experience	
	(1) Male	(2) Female	(3) > Median	(4) ≤ Median	(5) Yes	(6) No
Pre Comm. Leave-Out Mean	-0.07* (0.04)	-0.10** (0.04)	-0.09** (0.04)	-0.07* (0.04)	-0.11*** (0.04)	-0.05 (0.04)
Observations	2111	1154	1633	1632	1725	1540
R-Sq	0.497	0.613	0.585	0.481	0.544	0.530
Panel B: Leadership						
	Gender		Age		Previous Experience	
	(1) Male	(2) Female	(3) > Median	(4) ≤ Median	(5) Yes	(6) No
Pre Lead Leave-Out Mean	-0.03 (0.04)	-0.16*** (0.04)	-0.10*** (0.03)	-0.04 (0.04)	-0.04 (0.04)	-0.10** (0.04)
Observations	2124	1162	1647	1639	1737	1549
R-Sq	0.480	0.549	0.537	0.460	0.509	0.496
Panel C: Attitude						
	Gender		Age		Previous Experience	
	(1) Male	(2) Female	(3) > Median	(4) ≤ Median	(5) Yes	(6) No
Pre Att. Leave-Out Mean	0.01 (0.04)	-0.03 (0.05)	-0.01 (0.04)	-0.00 (0.05)	0.01 (0.04)	-0.00 (0.05)
Observations	2124	1154	1640	1638	1733	1545
R-Sq	0.451	0.504	0.494	0.440	0.474	0.462

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual composite communication and the independent variable is pre-course peer average composite communication. In Panel B, the dependent variable is post-course individual composite leadership and the independent variable is pre-course peer average composite leadership. In Panel C, the dependent variable is post-course individual composite attitude and the independent variable is pre-course peer average composite attitude. All columns control for the individual's pre-course ability for the outcome of interest in that panel. Additional controls in all columns include age (in days), race indicators, and previous experience. All regressions also include course indicators, section year, and section month fixed effects. The median age is 19.58 years.

References

- Abbott, B., Gallipoli, G., Meghir, C., & Violante, G. L. (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, *127*(6), 2569–2624.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, *91*(5), 1369–1401.
- Aleman, X. (2020). *Startup fundraising is the most tangible gender gap. How can we overcome it?* <https://techcrunch.com/2020/11/09/startup-fundraising-is-the-most-tangible-gender-gap-how-can-we-overcome-it/#:~:text=Gender%20equality%20in%20VC%20makes%20more%20business%20sense&text=In%20fact%2C%20a%20study%20by,%25%20from%20men%2Dfounded%20companies>. (Accessed: March 2021)
- Allison, P. D., & Waterman, R. P. (2002). Fixed-effects negative binomial regression models. *Sociological Methodology*, *32*(1), 247–265.
- Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, *1*(1), 136–63.
- Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455.
- Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspec-*

- tives*, 24(2), 3–30.
- Angrist, J. D., & Rokkanen, M. (2015). Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff. *Journal of the American Statistical Association*, 110(512), 1331–1344.
- Archibald, R. B., & Feldman, D. H. (2008). Explaining increases in higher education costs. *The Journal of Higher Education*, 79(3), 268–295.
- Baker, M., Gruber, J., & Milligan, K. (2015). *Non-cognitive deficits and young adult outcomes: The long-run impacts of a universal child care program* (Tech. Rep.). National Bureau of Economic Research.
- Balart, P., Oosterveen, M., & Webbink, D. (2018). Test scores, noncognitive skills and economic growth. *Economics of Education Review*, 63, 134–153.
- Baum, S., & Ma, J. (2007). *Trends in college pricing 2007*. The College Board Trends in Higher Education Series. Washington, DC..
- Beattie, G., Laliberté, J.-W. P., Michaud-Leclerc, C., & Oreopoulos, P. (2019). What sets college thrivers and divers apart? A contrast in study habits, attitudes, and mental health. *Economics Letters*, 178, 50–53.
- Bertanha, M., & Imbens, G. W. (2019). External validity in fuzzy regression discontinuity designs. *Journal of Business & Economic Statistics*, 1–39.
- Bertrand, M., & Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*, 5(1), 32–64.
- Biderman, C., De Mello, J. M., & Schneider, A. (2010). Dry laws and homicides: Evidence from the São Paulo metropolitan area. *The Economic Journal*, 120(543), 157–182.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics*, 93(2), 455–467.
- Bloom, H. S. (2012). Modern regression discontinuity analysis. *Journal of Research on Educational Effectiveness*, 5(1), 43–82.
- Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., & Eccles, J. (2006). Empathy or antipathy?

- The impact of diversity. *American Economic Review*, 96(5), 1890–1905.
- Borghans, L., Meijers, H., & Ter Weel, B. (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, 46(1), 2–12.
- Bormann, C. A., & Stone, M. H. (2001). The effects of eliminating alcohol in a college stadium: The Folsom Field beer ban. *Journal of American College Health*, 50(2), 81–88.
- Bound, J., & Turner, S. (2007). Cohort crowding: How resources affect collegiate attainment. *Journal of Public Economics*, 91(5-6), 877–899.
- Bowen, H. R. (1980). The costs of higher education: How much do colleges and universities spend per student and how much should they spend?
- Boyes, W. J., & Faith, R. L. (1993). Temporal regulation and intertemporal substitution: The effect of banning alcohol at college football games. *Public Choice*, 77(3), 595–609.
- Bruce, D. J., & Carruthers, C. K. (2014). Jackpot? The impact of lottery scholarships on enrollment in Tennessee. *Journal of Urban Economics*, 81, 30–44.
- Buckles, K. S., & Hungerman, D. M. (2013). Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics*, 95(3), 711–724.
- Bureau of the Census. (1998). *First findings from the EQW National Employer Survey*. EQW Catalog RE01. (Accessed: July 2020)
- Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2), 372–404.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, 101(3), 442–451.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), 2295–2326.
- Cameron, A. C., & Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372.
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

- Canaan, S. (2020). The long-run effects of reducing early school tracking. *Journal of Public Economics*, 187, 104206.
- Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy*, 100(1), 1–40.
- Carnevale, A. P., Cheah, B., & Hanson, A. R. (2015). *The economic value of college majors* (Tech. Rep.). Center on Education and the Workforce.
- Carpenter, C., & Dobkin, C. (2010). Alcohol Regulation and Crime. In *Controlling crime: Strategies and tradeoffs* (pp. 291–329). University of Chicago Press.
- Carpenter, C., & Dobkin, C. (2015). The minimum legal drinking age and crime. *Review of Economics and Statistics*, 97(2), 521–524.
- Carpenter, C., & Dobkin, C. (2017). The minimum legal drinking age and morbidity in the United States. *Review of Economics and Statistics*, 99(1), 95–104.
- Carrell, S. E., Hoekstra, M., & West, J. E. (2011a). Does drinking impair college performance? Evidence from a regression discontinuity approach. *Journal of Public Economics*, 95(1-2), 54–62.
- Carrell, S. E., Hoekstra, M., & West, J. E. (2011b). Is poor fitness contagious?: Evidence from randomly assigned friends. *Journal of Public Economics*, 95(7-8), 657–663.
- Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, 81(3), 855–882.
- Carruthers, C. K., & Özek, U. (2016). Losing HOPE: Financial aid and the line between college and work. *Economics of Education Review*, 53, 1–15.
- Carter, E. W., Sisco, L. G., Chung, Y.-C., & Stanton-Chapman, T. L. (2010). Peer interactions of students with intellectual disabilities and/or autism: A map of the intervention literature. *Research and Practice for Persons with Severe Disabilities*, 35(3-4), 63–79.
- Cattaneo, M. D., Crump, R. K., Farrell, M. H., & Feng, Y. (2019). On binscatter. *arXiv preprint arXiv:1902.09608*.

- Cattaneo, M. D., Frandsen, B. R., & Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the us senate. *Journal of Causal Inference*, 3(1), 1–24.
- Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.
- Cattaneo, M. D., & Jansson, M. (2018). Kernel-based semiparametric estimators: small bandwidth asymptotics and bootstrap consistency. *Econometrica*, 86(3), 955–995.
- Cattaneo, M. D., Jansson, M., & Ma, X. (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 1–7.
- Cellini, S. R., Ferreira, F., & Rothstein, J. (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, 125(1), 215–261.
- Chiteji, N. (2010). Time preference, noncognitive skills and well being across the life course: Do noncognitive skills encourage healthy behavior? *American Economic Review*, 100(2), 200–204.
- Clark, D., Gill, D., Prowse, V., & Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, 102(4), 648–663.
- Clark, D., & Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, 103(6), 2087–2120.
- Cohodes, S. R., & Goodman, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts’ adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6(4), 251–85.
- Cohodes, S. R., Grossman, D. S., Kleiner, S. A., & Lovenheim, M. F. (2016). The effect of child health insurance access on schooling: Evidence from public insurance expansions. *Journal of Human Resources*, 51(3), 727–759.
- College Board. (2019). *Average aid per student over time*. <https://research.collegeboard.org/insights/2019/08/average-aid-per-student-over-time>

[.collegeboard.org/trends/student-aid/figures-tables/average-aid-student-over-time](https://collegeboard.org/trends/student-aid/figures-tables/average-aid-student-over-time).

- Cook, T. D. (2008). “Waiting for life to arrive”: A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, 142(2), 636–654.
- Cornwell, C., & Mustard, D. B. (2007). Merit-based college scholarships and car sales. *Education Finance and Policy*, 2(2), 133–151.
- Cornwell, C., Mustard, D. B., & Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from georgia’s hope program. *Journal of Labor Economics*, 24(4), 761–786.
- Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student responses to merit scholarship retention rules. *Journal of Human Resources*, 40(4), 895–917.
- Cowan, B. W. (2016). Testing for educational credit constraints using heterogeneity in individual time preferences. *Journal of Labor Economics*, 34(2), 363–402.
- Cowan, B. W., & White, D. R. (2015). The effects of merit-based financial aid on drinking in college. *Journal of Health Economics*, 44, 137–149.
- Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2), 31–47.
- Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1, 697–812.
- Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3), 883–931.
- Cunningham, S., & Shah, M. (2018). Decriminalizing indoor prostitution: Implications for sexual violence and public health. *The Review of Economic Studies*, 85(3), 1683–1715.
- Dell, M., & Querubin, P. (2018). Nation building through foreign intervention: Evidence from discontinuities in military strategies. *The Quarterly Journal of Economics*, 133(2), 701–764.
- Deloitte. (2018). *Women in the boardroom: A global perspective*. <https://www2.deloitte.com/global/en/pages/risk/articles/women-in-the>

[-boardroom-global-perspective.html](#). (Accessed: March 2021)

- Deming, D., & Dynarski, S. (2008). The lengthening of childhood. *Journal of Economic Perspectives*, 22(3), 71–92.
- Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4), 1593–1640.
- Denning, J. T. (2019). Born under a lucky star? Financial aid, college completion, labor supply, and credit constraints. *Journal of Human Resources*, 54(3), 760–784.
- Dong, Y. (2018). Alternative assumptions to identify LATE in fuzzy regression discontinuity designs. *Oxford Bulletin of Economics and Statistics*, 80(5), 1020–1027.
- Dong, Y., & Lewbel, A. (2015). Identifying the effect of changing the policy threshold in regression discontinuity models. *Review of Economics and Statistics*, 97(5), 1081–1092.
- Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, 101(6), 2616–52.
- Dynarski, S. (2004). The new merit aid. In *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 63–100). University of Chicago Press.
- Dynarski, S., & Scott-Clayton, J. (2013). *Financial aid policy: Lessons from research* (Tech. Rep.). National Bureau of Economic Research.
- Edwards, H. (2020). *Official ACT to SAT (new 1600 and old 2400) conversion charts*. <https://blog.prepscholar.com/act-to-sat-conversion>.
- Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, 35(2), 387–428.
- Figlio, D., & Özek, U. (2020). An extra year to learn English? Early grade retention and the human capital development of English learners. *Journal of Public Economics*, 186, 104184.
- Fisher, R. A. (1935). *The design of experiments*. Oliver & Boyd.
- Frandsen, B. R. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. In *Regression discontinuity designs*. Emerald Publishing Limited.

- Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4), 191–210.
- Goldstein, P. J. (1985). The drugs/violence nexus: A tripartite conceptual framework. *Journal of Drug Issues*, 15(4), 493–506.
- Goodman, J. (2008). Who merits financial aid? *Journal of Public Economics*, 92(10-11), 2121–2131.
- Grembi, V., Nannicini, T., & Troiano, U. (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics*, 1–30.
- Grönqvist, H., & Niknami, S. (2014). Alcohol availability and crime: Lessons from liberalized weekend sales restrictions. *Journal of Urban Economics*, 81, 77–84.
- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201–209.
- Hahn, R. A., Kuzara, J. L., Elder, R., Brewer, R., Chattopadhyay, S., Fielding, J., ... Lawrence, B. (2010). Effectiveness of policies restricting hours of alcohol sales in preventing excessive alcohol consumption and related harms. *American Journal of Preventive Medicine*, 39(6), 590–604.
- Hayes, D. (2019). *College Football Programs That Sell In-Stadium Beer & Alcohol at Games in 2019 NCAAF Season*. <https://www.fanduel.com/theduel/posts/college-football-programs-that-sell-in-stadium-beer-alcohol-at-games-in-2019-ncaaf-season-01dje257jj5d>. (Accessed: June 2020)
- Heaton, P. (2012). Sunday liquor laws and crime. *Journal of Public Economics*, 96(1), 42–52.
- Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, 91(2), 145–149.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.
- Heller, D. E., & Marin, P. (2004). State merit scholarship programs and racial inequality. *Civil*

Rights Project at Harvard University (The).

- Hendren, N., & Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3), 1209–1318.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960.
- Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, 4(1), 93–106.
- Hoxby, C. M., & Weingarth, G. (2006). *Taking race out of the equation: School reassignment and the structure of peer effects*. (Working Paper)
- Hsu, Y.-C., & Shen, S. (2019). Testing treatment effect heterogeneity in regression discontinuity designs. *Journal of Econometrics*, 208(2), 468–486.
- Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3), 933–959.
- Jia, N. (2019). Heterogeneous effects of merit scholarships: Do program features matter? *Applied Economics*, 51(27), 2963–2979.
- Jones, T. R., & Kofoed, M. S. (2020). Do peers influence occupational preferences? Evidence from randomly-assigned peer groups at West Point. *Journal of Public Economics*, 184, 104154.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.
- Knox, T. (2016). *Texas, West Virginia and a Big Ten school have sold more beer than Ohio State*. <https://www.bizjournals.com/columbus/news/2016/10/14/texas-west-virginia-and-a-big-ten-school-have-sold.html>. (Accessed: June 2020)
- Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, 108(8), 2277–2304.
- Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. In *Handbook of the economics of education* (Vol. 5, pp. 1–74). Elsevier.

- Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics*, 116(3), 777–803.
- Lee, D. S. (2008). Randomized experiments from non-random selection in US House elections. *Journal of Econometrics*, 142(2), 675–697.
- Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2), 655–674.
- Lemieux, T., & Milligan, K. (2008). Incentive effects of social assistance: A regression discontinuity approach. *Journal of Econometrics*, 142(2), 807–828.
- Levitt, S. D. (1995). *Using electoral cycles in police hiring to estimate the effect of police on crime* (Tech. Rep.). National Bureau of Economic Research.
- Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, 8(4), 183–219.
- Li, Q., Lu, X., & Ullah, A. (2003). Multivariate local polynomial regression for estimating average derivatives. *Journal of Nonparametric Statistics*, 15(4-5), 607–624.
- Lindo, J. M., Siminski, P., & Swensen, I. D. (2018). College party culture and sexual assault. *American Economic Journal: Applied Economics*, 10(1), 236–65.
- Lindo, J. M., Siminski, P., & Yerokhin, O. (2016). Breaking the link between legal access to alcohol and motor vehicle accidents: Evidence from New South Wales. *Health Economics*, 25(7), 908–928.
- Lindo, J. M., Swensen, I. D., & Waddell, G. R. (2013). Alcohol and student performance: Estimating the effect of legal access. *Journal of Health Economics*, 32(1), 22–32.
- List, J. A., Momeni, F., & Zenou, Y. (2020). *The social side of early human capital formation: Using a field experiment to estimate the causal impact of neighborhoods* (Tech. Rep.). National Bureau of Economic Research.
- Lleras-Muney, A. (2002). Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939. *The Journal of Law and Economics*, 45(2), 401–435.

- Lleras-Muney, A., & Shertzer, A. (2015). Did the americanization movement succeed? An evaluation of the effect of English-only and compulsory schooling laws on immigrants. *American Economic Journal: Economic Policy*, 7(3), 258–90.
- MacDonald, J. M., Klick, J., & Grunwald, B. (2016). The effect of private police on crime: evidence from a geographic regression discontinuity design. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(3), 831–846.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3), 531–542.
- Marcus, J., & Siedler, T. (2015). Reducing binge drinking? The effect of a ban on late-night off-premise alcohol sales on alcohol-related hospital stays in Germany. *Journal of Public Economics*, 123, 55–77.
- Mas, A. (2006). Pay, reference points, and police performance. *The Quarterly Journal of Economics*, 121(3), 783–821.
- Matta, R., Ribas, R. P., Sampaio, B., & Sampaio, G. R. (2016). The effect of age at school entry on college admission and earnings: A regression-discontinuity approach. *IZA Journal of Labor Economics*, 5(1), 9.
- McCrary, J. (2007). The effect of court-ordered hiring quotas on the composition and quality of police. *American Economic Review*, 97(1), 318–353.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698–714.
- Menaker, B. E., & Chaney, B. H. (2014). College football game day stadium incidents: Policy and environmental effects on alcohol-related ejections and crime. *Journal of Policy Research in Tourism, Leisure and Events*, 6(2), 119–134.
- Miller, S., & Wherry, L. R. (2019). The long-term effects of early life medicaid coverage. *Journal of Human Resources*, 54(3), 785–824.
- Nakaguma, M. Y., & Restrepo, B. J. (2018). Restricting access to alcohol and public health: Evidence from electoral dry laws in Brazil. *Health Economics*, 27(1), 141–156.

- Neidell, M., & Waldfogel, J. (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics*, 92(3), 562–576.
- NIAAA. (2015). *Planning alcohol interventions using niaaa's collegeaim (alcohol intervention matrix)*. National Institute on Alcohol Abuse and Alcoholism Washington, DC.
- Ortagus, J., Kelchen, R., Rosigner, K., & Voorhees, N. (2018). Performance-based funding in American higher education: A systematic synthesis of the intended and unintended consequences. *Educational Evaluation and Policy Analysis*, 53(2), 299–329.
- Peña, P. A. (2017). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review*, 56, 152–176.
- Penn Live. (2011). *Beer sales at West Virginia University football games approved by board*. https://www.pennlive.com/patriotnewssports/2011/06/beer_sales_at_west_virginia_un.html. (Accessed: August 2020)
- Pettersson-Lidbom, P. (2012). Does the size of the legislature affect the size of government? Evidence from two natural experiments. *Journal of Public Economics*, 96(3-4), 269–278.
- Rea, D., & Burton, T. (2020). New evidence on the Heckman Curve. *Journal of Economic Surveys*, 34(2), 241–262.
- Rees, D. I., & Schnepel, K. T. (2009). College football games and crime. *Journal of Sports Economics*, 10(1), 68–87.
- Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, 116(2), 681–704.
- Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education* (Vol. 3, pp. 249–277). Elsevier.
- Sacks, J. J., Gonzales, K. R., Bouchery, E. E., Tomedi, L. E., & Brewer, R. D. (2015). 2010 national and state costs of excessive alcohol consumption. *American Journal of Preventive Medicine*, 49(5), e73–e79.
- Sackstein, S. (2019). *Why has the cost of college outpaced inflation?* <https://>

blogs.edweek.org/teachers/work_in_progress/2019/04/why_has_the_cost_of_college_ou.html. (Accessed: June 2020)

- Sanderson, A. R., & Siegfried, J. J. (2015). The case for paying college athletes. *Journal of Economic Perspectives*, 29(1), 115–38.
- Scott-Clayton, J. (2011). On money and motivation a quasi-experimental analysis of financial incentives for college achievement. *Journal of Human resources*, 46(3), 614–646.
- Scott-Clayton, J., & Zafar, B. (2019). Financial aid, debt management, and socioeconomic outcomes: Post-college effects of merit-based aid. *Journal of Public Economics*, 170, 68–82.
- Sibthorp, J., Paisley, K., & Gookin, J. (2007, January). Exploring participant development through adventure-based programming: A model from the national outdoor leadership school. *Leisure Sciences*, 29(1), 1–18.
- Sjoquist, D. L., & Winters, J. V. (2014). Merit aid and post-college retention in the state. *Journal of Urban Economics*, 80, 39–50.
- Sjoquist, D. L., & Winters, J. V. (2015a). The effect of Georgia’s HOPE scholarship on college major: a focus on stem. *IZA Journal of Labor Economics*, 4(1), 1–29.
- Sjoquist, D. L., & Winters, J. V. (2015b). State merit aid programs and college major: A focus on STEM. *Journal of Labor Economics*, 33(4), 973–1006.
- Sjoquist, D. L., & Winters, J. V. (2015c). State merit-based financial aid programs and college attainment. *Journal of Regional Science*, 55(3), 364–390.
- Smith, A. C. (2016). Spring forward at your own risk: Daylight saving time and fatal vehicle crashes. *American Economic Journal: Applied Economics*, 8(2), 65–91.
- Solis, A. (2017). Credit access and college enrollment. *Journal of Political Economy*, 125(2), 562–622.
- Takaku, R., & Yokoyama, I. (2021). What the covid-19 school closure left in its wake: Evidence from a regression discontinuity analysis in Japan. *Journal of Public Economics*, 195, 104364.
- Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative

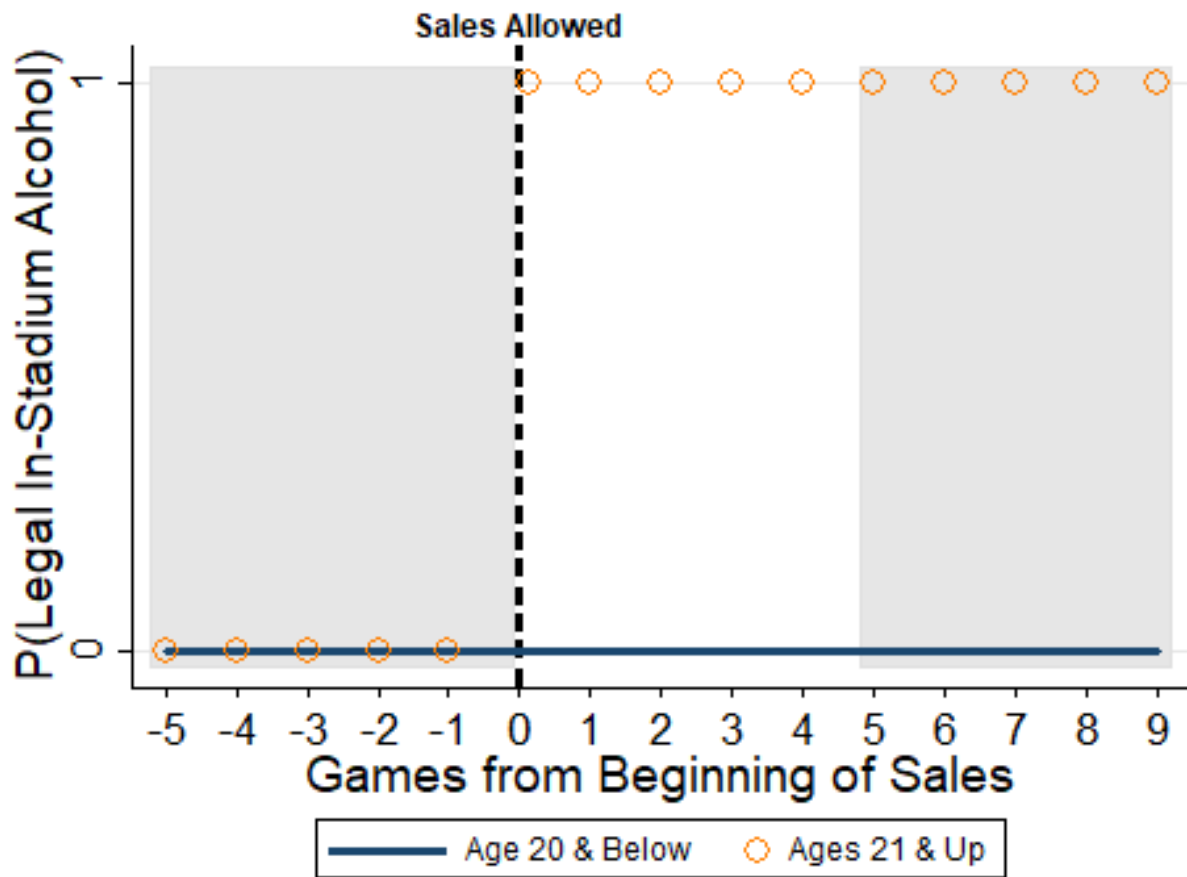
- to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), 309.
- Tracy, M. (2015). *Beer here! Beer here?* <https://www.nytimes.com/2015/10/09/sports/ncaafootball/beer-alcohol-college-football-west-virginia.html>. (Accessed: June 2020)
- Watkins, L., O'Reilly, M., Kuhn, M., Gevarter, C., Lancioni, G. E., Sigafos, J., & Lang, R. (2015). A review of peer-mediated social interaction interventions for students with autism in inclusive settings. *Journal of Autism and Developmental Disorders*, 45(4), 1070–1083.
- Weinberger, C. J. (2014). The increasing complementarity between cognitive and social skills. *Review of Economics and Statistics*, 96(4), 849–861.
- West Virginia Higher Education Policy Commission. (2014). *Final report of Promise scholarship ad-hoc advisory committee*. <http://www.wvhepc.edu/wp-content/uploads/2014/01/PROMISEfinalreport.pdf>.
- Whaley, M. (1985). The status of kindergarten: A survey of the states.
- Wherry, L. R., Miller, S., Kaestner, R., & Meyer, B. D. (2018). Childhood Medicaid coverage and later-life health care utilization. *Review of Economics and Statistics*, 100(2), 287–302.
- Whiting, K. (2020). *These are the top 10 job skills of tomorrow – and how long it takes to learn them*. <https://www.weforum.org/agenda/2020/10/top-10-work-skills-of-tomorrow-how-long-it-takes-to-learn-them/>. (Accessed: January 2021)
- Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1), 77–97.
- Zhang, L., Hu, S., Sun, L., & Pu, S. (2016). The effect of Florida's Bright Futures Program on college choice: A regression discontinuity approach. *The Journal of Higher Education*, 87(1), 115–146.

Appendix

A Essay 1

A.1 Additional Tables and Figures

Figure A.1: Time-Varying Difference in In-Stadium Legal Alcohol Availability by Age Group



Note: Gray shaded areas to non-shaded indicate beginning of new season. In the United States, the minimum legal drinking age is federally set to age 21, since the 1970's.

Table A.1: Colleges In Sample

Institution	First Sales Season	Agencies
Akron	2008-2009, 2012-2015	2
Bowling Green University	2009	2
Cincinnati	2011	1
Colorado State University	2008	2
Colorado-Boulder University	2014	3
Louisville University	2008	1
Memphis	2009	2
North Texas	2014	1
Toledo	2013	1
West Virginia University	2011	2
Western Kentucky University	2012	2
Air Force	Never	1
Arkansas	Never	2
Arkansas State University	Never	2
Boise State University	Never	2
Brigham Young University	Never	2
Central Michigan University	Never	2
Clemson	Never	2
Eastern Michigan University	Never*	2
Idaho University	Never	1
Iowa	Never	2
Iowa State University	Never	2
Kansas	Never	2
Kansas State	Never	2
Kentucky	Never	2
Marshall	Never	2
Michigan State University	Never	2
Michigan University	Never	2
Middle Tennessee University	Never	2
Ohio	Never	2
Ohio State University	Never	2
Oregon State University	Never	1
South Carolina University	Never	2
Tennessee	Never	2
Texas Christian University	Never	1
Texas Tech University	Never	1
Utah	Never	1
Utah State University	Never	2
Vanderbilt	Never	2
Virginia	Never	2
Virginia Polytechnic University	Never	2
Washington	Never	1
Washington State University	Never	2
Western Michigan University	Never	3

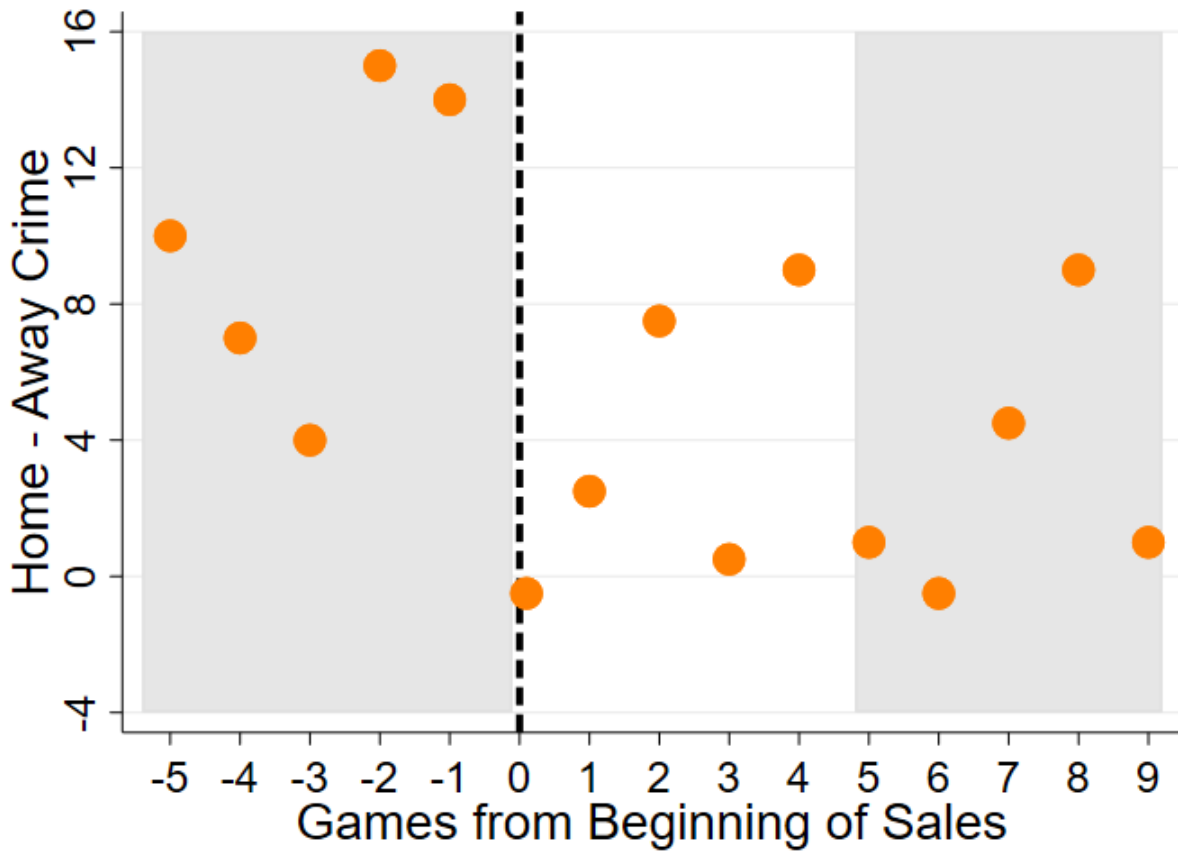
Source: Author reviewed internet for news regarding all teams in the NIBRS datasets in D-1 College Football. First Sales Season column is not necessarily the same year that sales were introduced, it is the years sales begin in the sample, meaning 2008 stands for a team that had sales in the stadium the entire sample period. Eastern Michigan had sales for one game in 2015.

Table A.2: Offense Descriptions

Code	Offense	Description
13A	Aggravated Assault	An unlawful attack on one person upon another wherein the offender uses a weapon or displays it in a threatening manner or the victim suffers obvious severe bodily injury involving apparent broken bones, loss of teeth, internal injury, severe laceration, or loss of consciousness
13B	Simple Assault	An unlawful physical attack by on one person upon another where neither the offender displays a weapon, nor the victim suffers obvious severe bodily injury involving apparent broken bones, loss of teeth, internal injury, severe laceration, or loss of consciousness
13C	Intimidation	To unlawfully place another person in reasonable fear of bodily harm through the use of threatening words and/or other conduct, but without displaying a weapon or subjecting the victim to actual physical attack
290	Vandalism	To willfully or maliciously destroy, damage, deface, or otherwise injure real or personal property without the consent of the owner or person having custody or control of it
11A	Forcible Rape	The carnal knowledge of a person, forcibly and/or against that persons will or against the persons will in instances where the victim is incapable of giving consent because of his/her temporary or permanent mental or physical incapacity
90C	Disorderly Conduct	Any behavior that tends to disturb the public or decorum, scandalize the community, or shock the public sense of morality
90D	Driving Under the Influence	Driving or operating a motor vehicle or common carrier while mentally or physically impaired as the result of consuming an alcoholic beverage or using a drug or narcotic
90E	Drunkenness	To drink alcoholic beverages to the extent that ones mental faculties and physical coordination are substantially impaired
90G	Liquor Law Violations	The violation of laws or ordinances prohibiting the manufacture, sale, purchase, transportation, possession, or use of alcoholic beverages

Note: Source is the National Incident-Based Reporting System (NIBRS) codebook.

Figure A.2: Trends in Crime, Home-Away



Notes: Dependent variable is home crime - away crime. Game 0 is the beginning of alcohol sales. Gray background indicates beginning of a new season.

Table A.3: Alcohol Sales Legalizations Effect on Total Crime at Home Games, With Exact Date Fixed Effects

Panel A: Without Away and Away Interaction	
	(1)
Home	0.19*** (0.06)
HomeXSales	-0.19** (0.09)
Observations	162898
Number of Agencies	79
Agency FE	X
Years	X
Months	X
Week of Yr	X
Day of Week	X
Holidays	X
Panel B: With Away and Away Interaction	
	(1)
Home	0.18*** (0.06)
HomeXSales	-0.19** (0.09)
Away	-0.01 (0.01)
AwayXSales	-0.01 (0.04)
Observations	162898
Number of Agencies	79
Agency FE	X
Years	X
Months	X
Week of Yr	X
Day of Week	X
Holidays	X

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors cluster-robust at the city-level in parentheses. All models estimated using Poisson conditional fixed effects models. Holidays are: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day. All models contain 2,062 exact date fixed effects.

Table A.4: Robustness to Model Type

	(1) Negative Binomial	(2) OLS
Home	0.72*** (0.11)	6.71*** (1.62)
HomeXSales	-0.52*** (0.19)	-5.82*** (1.99)
Observations	162898	162898

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All models estimated using conditional fixed effects models. Standard errors cluster-robust at the city-level in parentheses. All models include agency, year, and month fixed effects. All models use indicator variables for day of week and holidays including: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day. Column 1 uses a negative binomial model. Column 2 uses OLS.

Table A.5: Alcohol Sales Legalizations Effect on Different Crime Types

Panel A: Assault Types			
	(1) Aggravated	(2) Simple	(3) Intimidation
Home	0.02 (0.04)	0.12*** (0.03)	0.04 (0.03)
HomeXSales	-0.12*** (0.04)	-0.18*** (0.05)	-0.07** (0.04)
Panel B: Bad Behaviors			
	(1) Vandalism	(2) Rape	(3) Disorderly Conduct
Home	0.10*** (0.03)	0.11*** (0.04)	0.52*** (0.13)
HomeXSales	-0.03 (0.07)	-0.06 (0.13)	-0.50** (0.20)
Panel C: Liquor Crime Types			
	(1) Liquor Law Violations	(2) Drunkenness	(3) Driving Under the Influence
Home	1.19*** (0.19)	0.74*** (0.21)	0.11*** (0.04)
HomeXSales	-1.21*** (0.37)	-0.48 (0.31)	-0.05 (0.08)
Panel D: Falsification Crimes			
	(1) Grand Theft Auto	(2) Embezzlement	(3) Peeping Tom
Home	0.07** (0.03)	0.05 (0.08)	0.65 (0.54)
HomeXSales	-0.05 (0.04)	-0.14 (0.15)	0.31 (1.01)

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors cluster-robust at the city-level in parentheses. All models estimated using Poisson conditional fixed effects models. Holidays are: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day. All models include agency fixed effects, day of week indicators, year fixed effects, and month fixed effects.

Table A.6: Alcohol Sales Legalizations Effect on Total Crime at Home Games

Panel A: Alternative Dependent Variables					
	Game-Associated Crime			Alcohol Crimes	
	(1)	(2)	(3)	(4)	(5)
Home	0.12*** (0.04)	0.20*** (0.06)	0.35*** (0.10)	0.65*** (0.16)	1.18*** (0.17)
HomeXSales	-0.11** (0.06)	-0.24*** (0.09)	-0.39*** (0.13)	-0.63*** (0.23)	-0.49* (0.29)
Observations	162898	162898	162898	162898	162898
Number of Agencies	79	79	79	79	79
Drop Liquor Law Viol.	X	X	-	-	X
Game-Crime Subset	-	X	X	-	-
Alc. Suspected	-	-	-	X	X
Crimes Within 3 Hours of Game	-	-	-	-	X
Panel B: Alternative Subsamples/Control Groups					
	Placebo	Gamedays		Agency Type	
	(1)	(2)	(3)	(4)	(5)
Away	-0.04** (0.02)				
AwayXSales	-0.02 (0.03)				
Home		0.28*** (0.09)	0.45*** (0.12)	1.29*** (0.14)	0.09** (0.04)
HomeXSales		-0.24** (0.09)	-0.38*** (0.13)	-0.90*** (0.26)	-0.10* (0.06)
Observations	162898	7442	7442	65984	96914
Number of Agencies	79	79	79	32	47
Gamedays Only	-	X	X	-	-
Drop Liquor Law Viol.	-	X	-	-	-
University Agencies Only	-	-	-	X	-
City Agencies Only	-	-	-	-	X

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors cluster-robust at the city-level in parentheses. All models estimated using Poisson conditional fixed effects models. In Panel B and C, all models include agency, year, and month fixed effects, day of week indicators, and holiday indicators. Holidays are: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day. Panel B, column 1 drops liquor law violations. Panel B, column 2 includes only crimes in the following types: assault, rape, vandalism, drunkenness, dui. Panel B, column 3 adds liquor law violations back. Panel B, column 4 only includes crimes where alcohol was suspected For Type A crimes, this means an alcohol suspected indicator is present; for Type B crimes alcohol must be in the name (liquor law violation, drunkenness, or DUI) of being used by the offender. Panel B, column 5 only includes crimes where alcohol was suspected and the crime occurred within 3 hours of the game beginning or ending. Panel C columns 2 and 3 include only home and away games. Panel C, column 4 is only university agencies. Panel C, column 5 is only city/county agencies.

Table A.7: Effect of Alcohol Sales on Specific, Related Crime Types

	(1)	(2)
Home	0.10*** (0.03)	0.06*** (0.02)
HomeXSales	-0.11* (0.07)	-0.07** (0.03)
Observations	162898	162898
B Crimes Incl.	None	None
Location Types Incl.	Narrow	Broad

Note: All models estimated using Poisson conditional fixed effects models. Standard errors cluster-robust at the city-level in parentheses. All models include agency, year, and month fixed effects. All models use indicator variables for day of week and holidays including: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day. Panel A, column 1 only includes crimes that are committed in the following locations: bar/nightclub, street, liquor store, parking lot, school/college, and arena/stadium. Panel A, column 2 uses the same crime locations and adds: convenience stores, field/woods, grocery store, residence, restaurant, gas station, and gambling facility. All models include agency fixed effects, day of week indicators, year fixed effects, and month fixed effects.

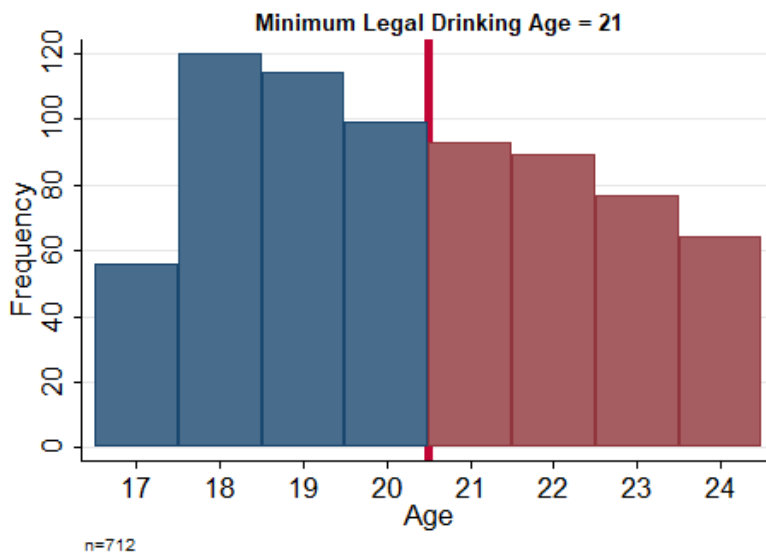
Table A.8: Investigating the Influence of Omitted Attendance and Law Enforcement on Effect of Sales on Crime

Panel A: Dependent Variable: Total Crime		
	(1)	(2)
Home	-0.01 (0.14)	0.19*** (0.06)
HomeXSales	-0.08 (0.08)	-0.19** (0.08)
Home Attendance (10000s)	0.04 (0.03)	
Home Attend. Deviation (10000s)		0.03 (0.04)
Observations	162898	162898
Panel B: Dependent Variable: Game-Alcohol Related Crimes		
	(1)	(2)
Home	0.13 (0.22)	0.35*** (0.10)
HomeXSales	-0.27** (0.13)	-0.39*** (0.14)
Home Attendance (10000s)	0.04 (0.05)	
Home Attend. Deviation (10000s)		0.08 (0.05)
Observations	162898	162898

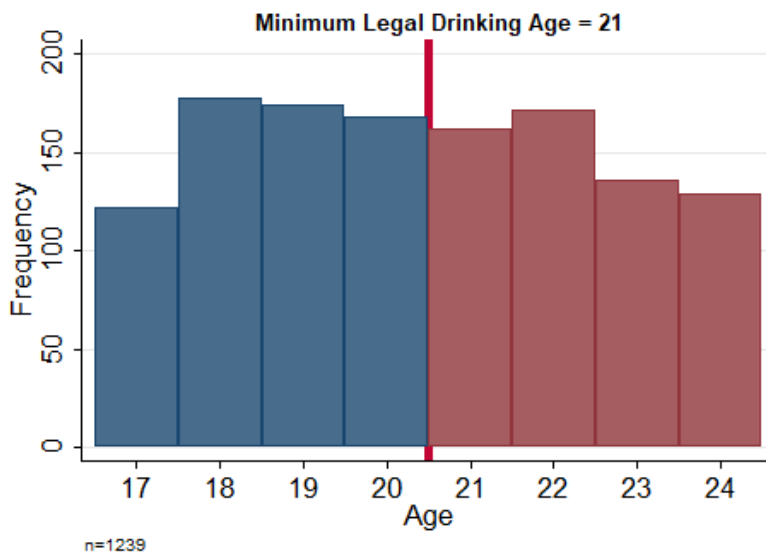
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors cluster-robust at the city-level in parentheses. All models include agency, year, month fixed effects. All models use indicator variables for day of week and holidays including: Thanksgiving, Halloween, Labor Day, Columbus Day, and Veterans Day. Panels A and B are estimated using Poisson conditional fixed effects models. In Panels A and B, column 1 includes attendance in 10,000's. Non-home game attendance is assumed to be 0. In Panels A and B, column 2 home game attendance is subtracted from the average within-sample attendance. In Panel B, the dependent variable is game/alcohol-related crime, including: assault, rape, vandalism, drunkenness, DUI, and liquor law violations.

Figure A.3: Examining If Age-Based Variation Feasibly Controls for Enforcement Bias; Density of Having a Report

(a): Pre-Sales Home Gameday Clear. Rate Density



(b): Post-Sales Home Gameday Clear. Rate Density

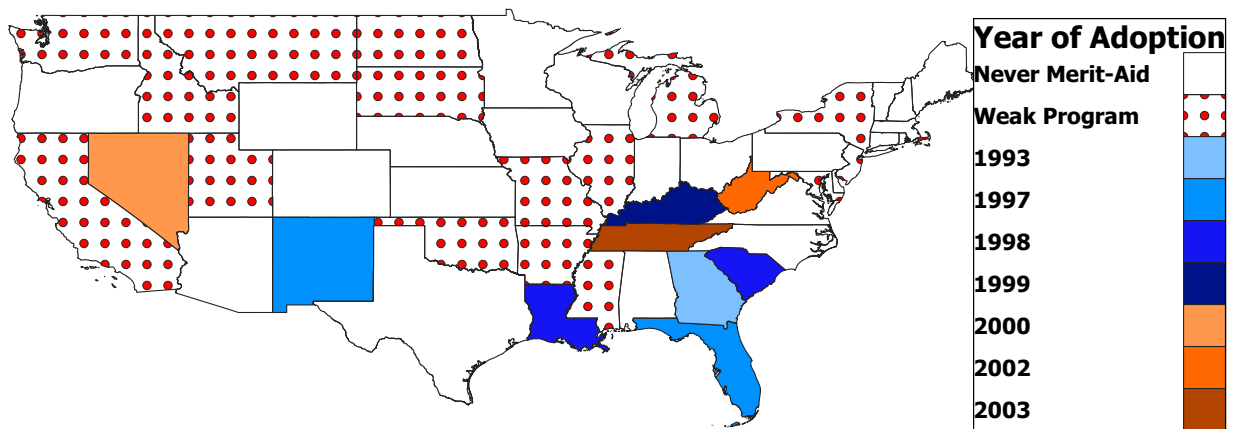


Note: Panel A and B show the likelihood of having at least 1 report on gameday by age. At age 18, individuals become a legal adult.

B Essay 2

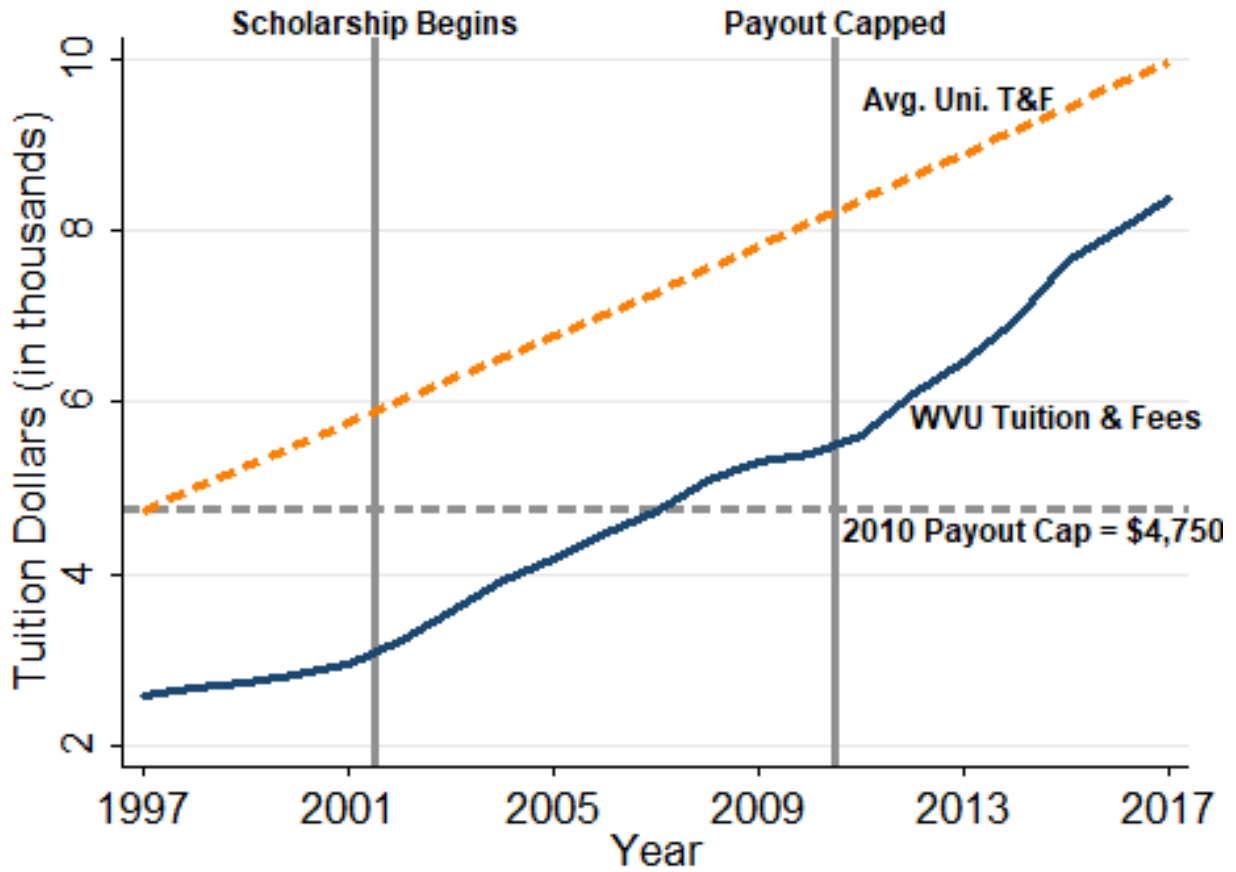
B.1 Additional Tables and Figures

Figure B.1: Merit-Aid Adoption by State and Time



Note: Data comes from [Dynarski \(2004\)](#), [Heller and Marin \(2004\)](#), and various state agencies. Weak program means the program does not cover many students or is not generous. Weak programs date of beginning not shown.

Figure B.2: Trends in Cost of College and West Virginia Promise Scholarship Program



Note: West Virginia University tuition and fees come from West Virginia University website. Average university tuition and fees comes from College Board. There are 3 observations, in 1997, 2007, and 2017. It is linearly interpolated for other years. Data on WVU T&F from <https://institutionalresearch.wvu.edu/institutional-research/statistical-information/tuition-and-fees>.

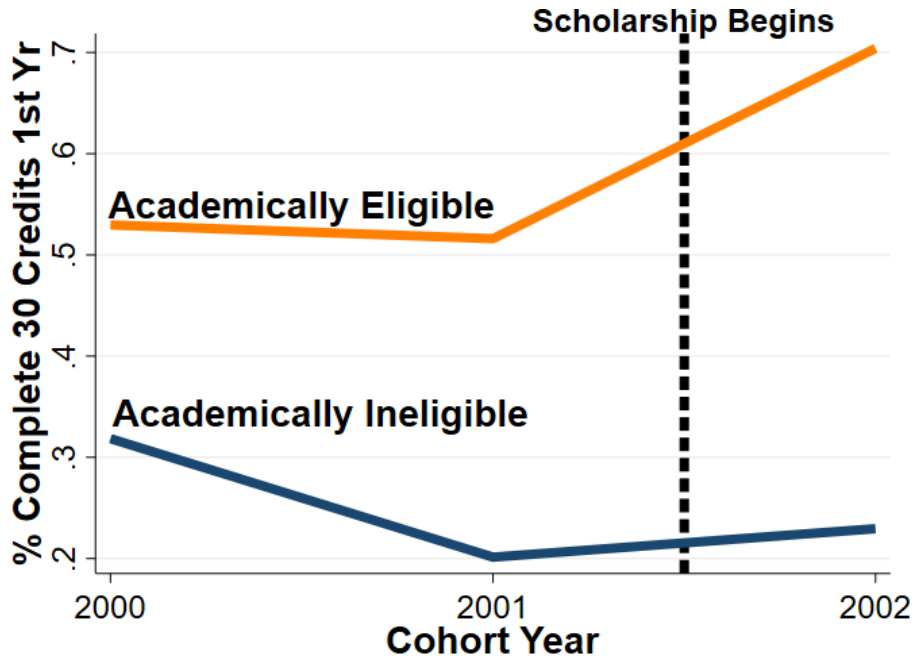
Table B.1: Descriptive Statistics by Eligibility

Panel A: Acad Elig., Born Day ≥ 0						
	Mean	Median	Std Dev	Min	Max	Count
Freshman Credits	28.01	30	8.13	0	49	1601
Freshman Renewal Credits	0.62	1	0.49	0	1	1601
Exceed Freshman Renew GPA	0.65	1	0.48	0	1	1601
Scholarship Years	2.07	2	1.71	0	4	1601
Panel B: Acad Elig., Born Day < 0						
	Mean	Median	Std Dev	Min	Max	Count
Freshman Credits	26.78	29	8.53	0	53	1402
Freshman Renewal Credits	0.46	0	0.50	0	1	1402
Exceed Freshman Renew GPA	0.59	1	0.49	0	1	1402
Scholarship Years	0.10	0	0.60	0	4	1402
Panel C: Not Acad Elig., Born Day ≥ 0						
	Mean	Median	Std Dev	Min	Max	Count
Freshman Credits	19.86	22	10.34	0	38	273
Freshman Renewal Credits	0.20	0	0.40	0	1	273
Exceed Freshman Renew GPA	0.24	0	0.43	0	1	273
Scholarship Years	0.13	0	0.63	0	4	273
Panel D: Not Acad Elig., Born Day < 0						
	Mean	Median	Std Dev	Min	Max	Count
Freshman Credits	18.36	22	10.63	0	39	310
Freshman Renewal Credits	0.13	0	0.33	0	1	310
Exceed Freshman Renew GPA	0.22	0	0.41	0	1	310
Scholarship Years	0.00	0	0.06	0	1	310

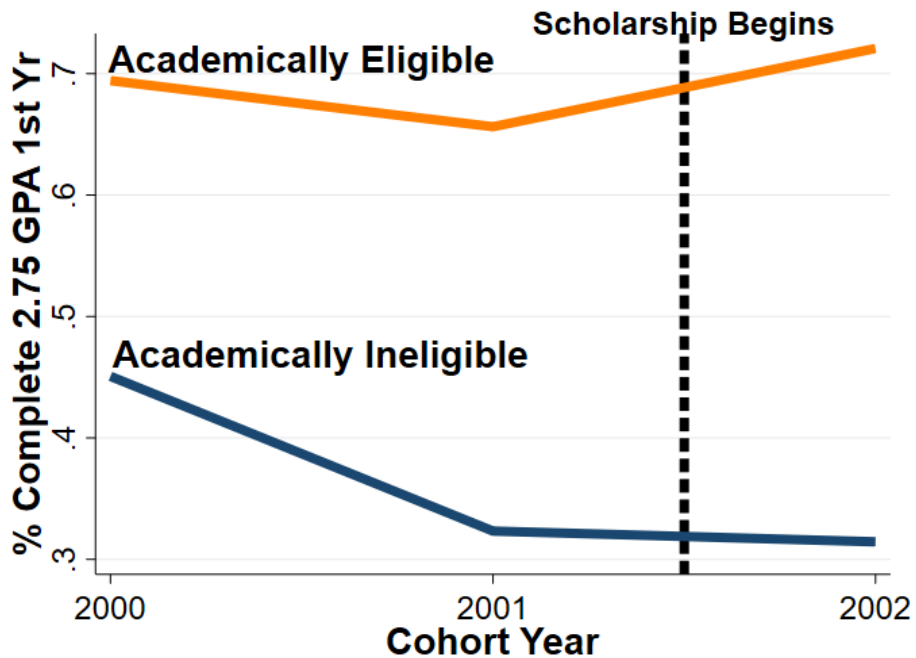
Note: Student-level observations. In-state students at WVU. First-time freshman only. Only cohorts 2001 and 2002. Freshman renewal GPA = 2.75. Freshman renewal credits = 30. Academic eligibility obtained by (HS GPA > 3) AND (ACT Composite > 20 OR SAT Composite > 1000). Temporal eligibility if born on or after cutoff date. Panel A includes academically eligible and temporally eligible students. Panel B includes academically eligible, temporally ineligible. Panel C includes non-academically eligible, temporally eligible. Panel D includes academically ineligible and temporally ineligible students.

Figure B.3: 1st Year Renewal Likelihood by Cohort

(a): First Year Credit Renewal

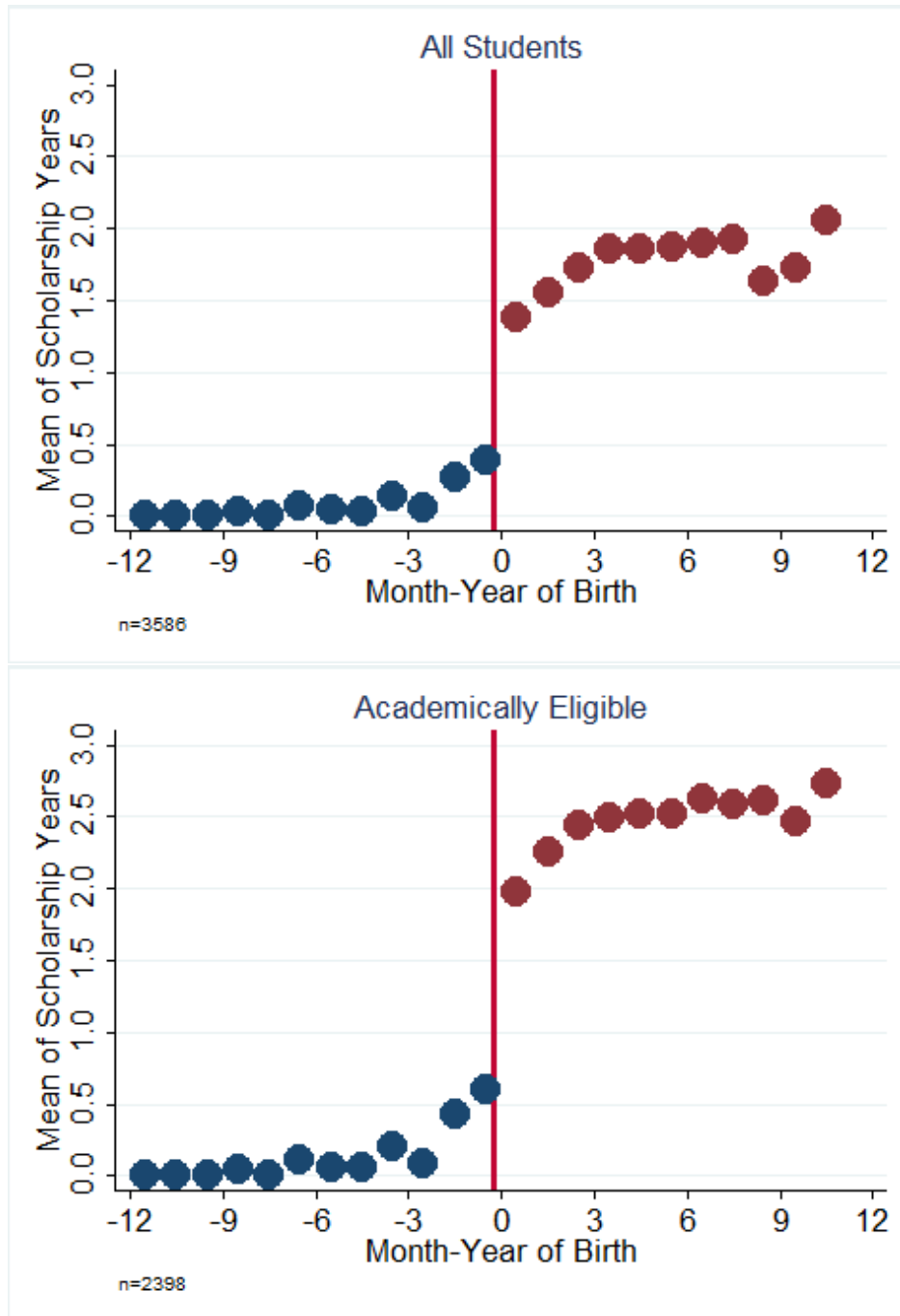


(b): First Year GPA Renewal



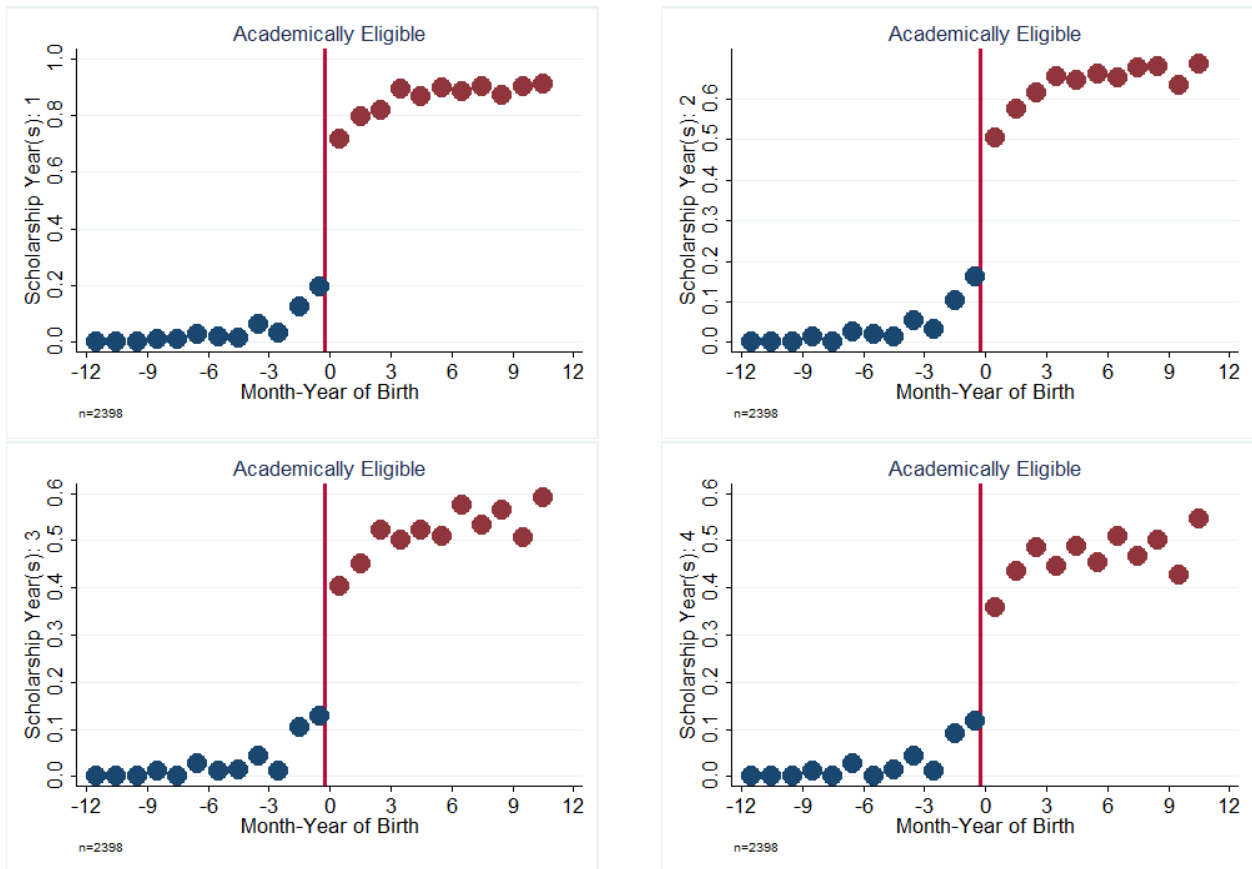
Note: A shows the percent of students who exceed the number of completed credits (30) required for scholarship renewal, once the scholarship begins, by cohort and academic eligibility. B shows the percent of students who exceed the GPA requirement required for scholarship renewal, once the scholarship begins, by cohort and academic eligibility.

Figure B.4: Effect of Birth Month on Years of Scholarship



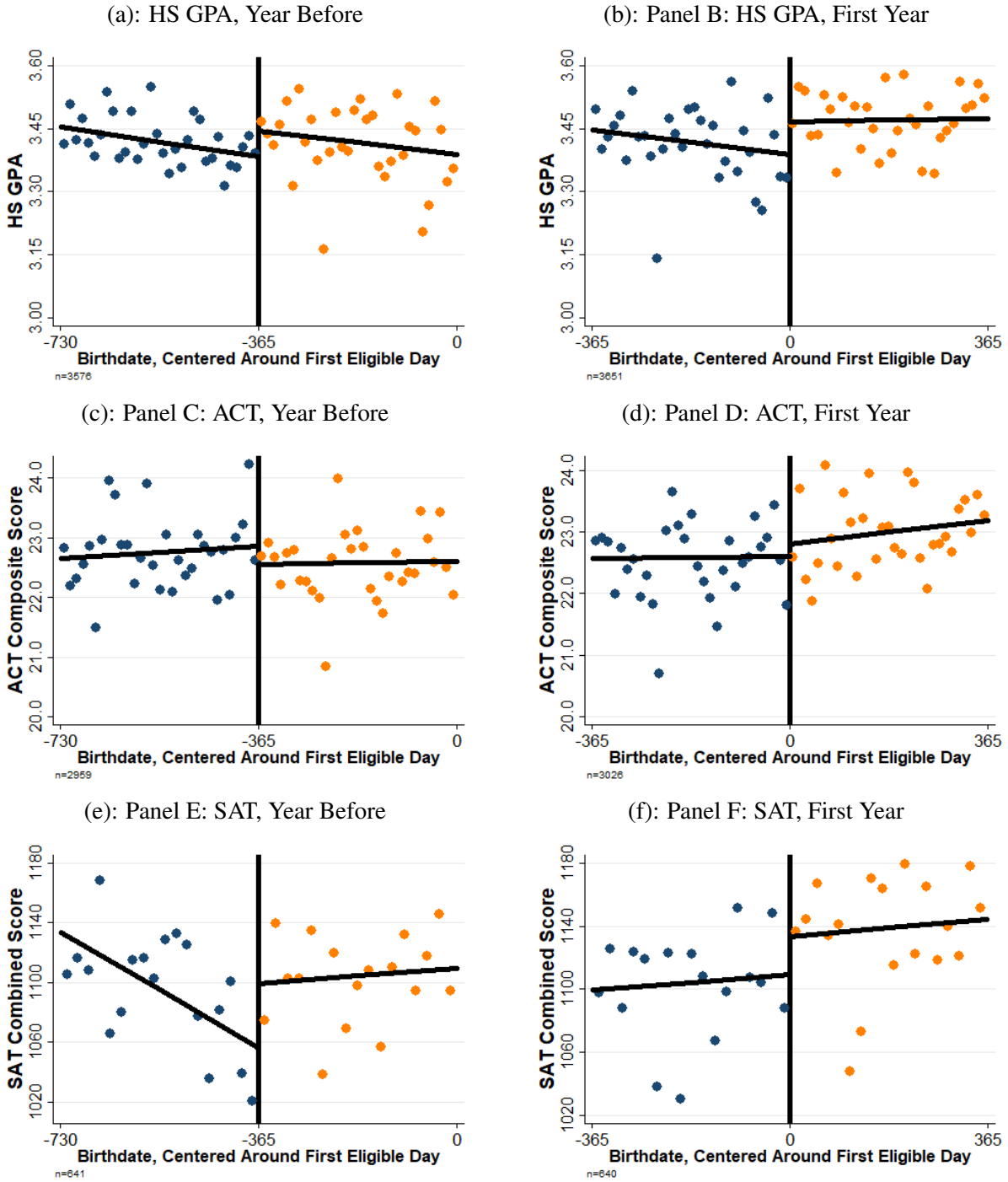
Note: Academic eligibility determined by high school GPA and ACT score.

Figure B.5: Relationship between Birth Date and Years of Scholarship Received, Academically Eligible



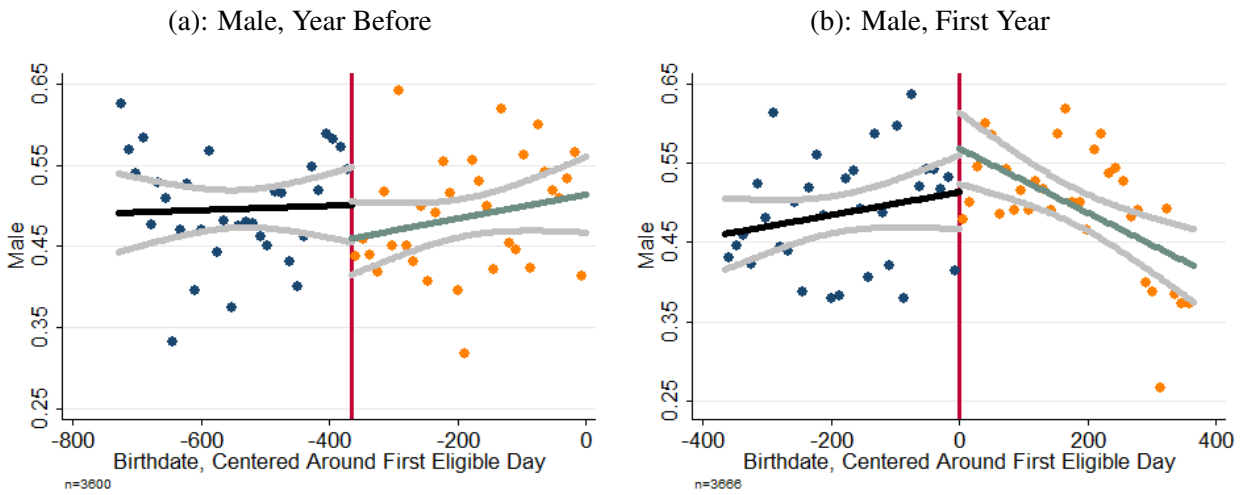
Note: Each successive graph uses a more restrictive criteria for the likelihood of scholarship. The y-axis measures likelihood of receiving scholarship for 1-4 years, 4 being the maximum number of years allowed by the policy.

Figure B.6: Pre-Treatment Covariate Continuity at Date Cutoff



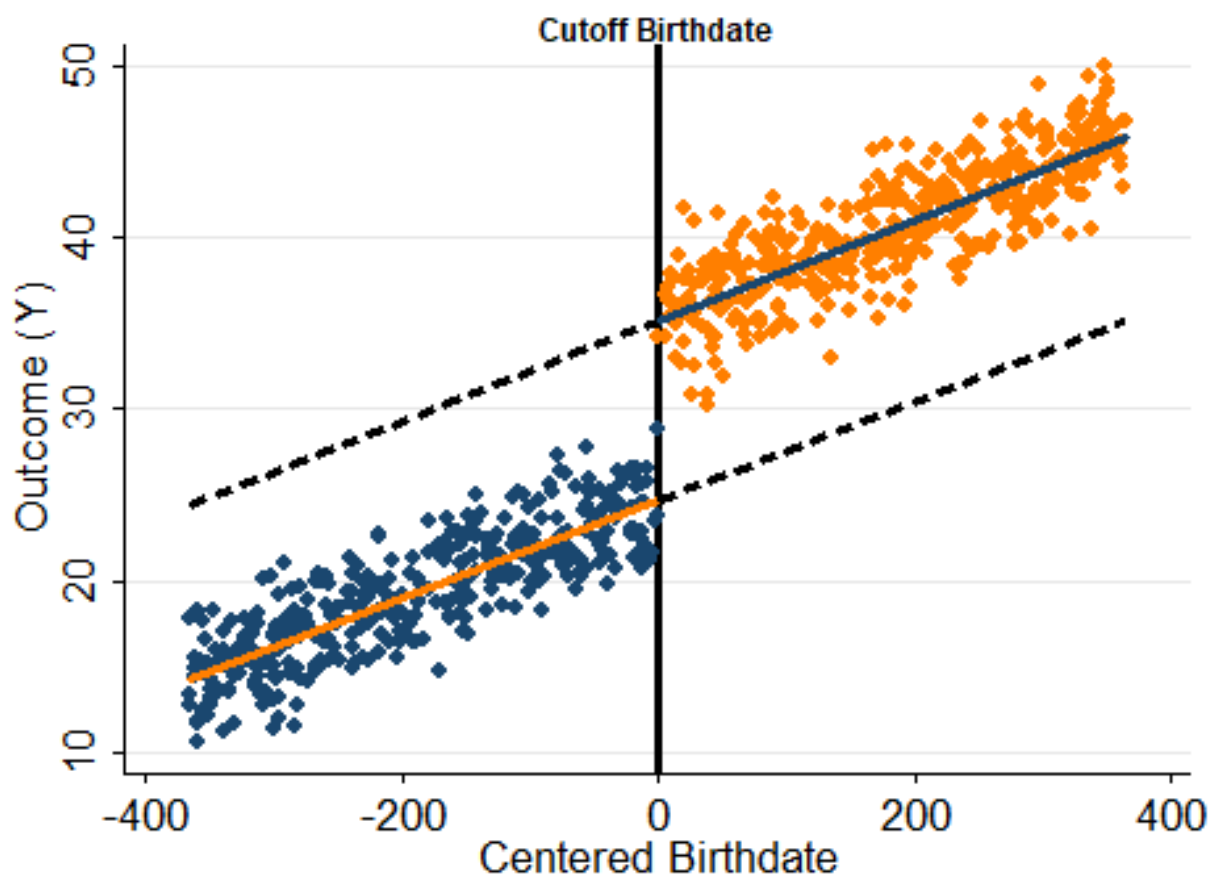
Note: Red line represents cutoff birthdate. Panels A, C, and E use the year prior to the scholarship being available cutoff, 365 days before 0. Panels B, D, and F use the year where the cutoff determines eligibility. Points are means conditional on point being within a bandwidth (bw) of h of a given birthdate. Plot A and B $bw = 11.375$. Plots C and D $bw = 11.74$. Plots E and F $bw = 20.11$. Bandwidths chosen as default histogram option. Specifically: number of bins chosen as: $\min(\sqrt{N}, \frac{10 \ln(N)}{\ln(10)})$, where N is number of observations.

Figure B.7: Tests of Gender Continuity at Date Cutoff



Note: Red line represents cutoff birthdate. Panels A, C, and E use the year prior to the scholarship being available cutoff, 365 days before 0. Panels B, D, and F use the year where the cutoff determines eligibility. Points are means conditional on point being within a bandwidth (bw) of h of a given birthdate. Plot A and B $bw = 11.375$. Plots C and D $bw = 11.74$. Plots E and F $bw = 20.11$. Bandwidths chosen as default histogram option. Specifically: number of bins chosen as: $\min(\sqrt{N}, \frac{10 \ln(N)}{\ln(10)})$, where N is number of observations.

Figure B.8: Regression Discontinuity Identification Assumption: Local Smoothness of Potential Outcomes

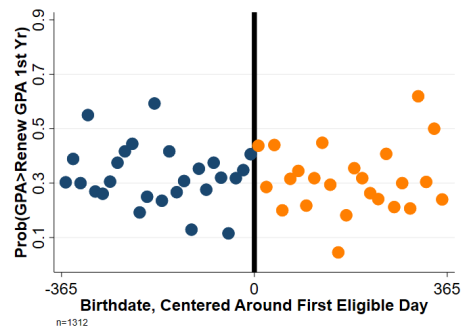
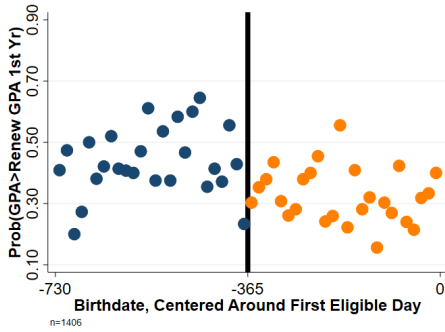


Note: Data is author-simulated. Picture depicts the assumption regression discontinuity, local smoothness, makes to estimate causal effects. Under local smoothness of potential outcomes, units just above and below make a valid counterfactual for one another.

Figure B.9: Reduced Form Effect of Birthdate on Likelihood of Exceeding Renewal Thresholds in Freshman Year, Academically Ineligible Students

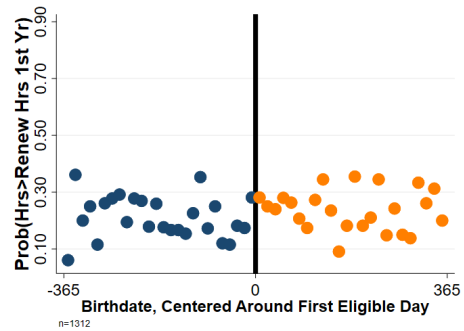
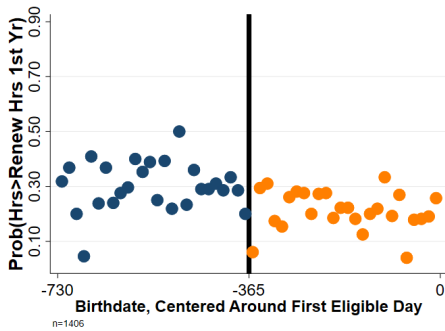
(a): Exceed Renewal GPA, Pre-Scholarship Cutoff

(b): Exceed Renewal GPA, Scholarship Cutoff



(c): Exceed Renewal Credits, Pre-Scholarship Cutoff

(d): Exceed Renewal Credits, Scholarship Cutoff



Note: Red line represents cutoff birthdate. Panels A & C use the year prior to the scholarship being available cutoff, 365 days before 0. Panels B & D use the year where the cutoff determines scholarship eligibility. Points are means conditional on point being within a bandwidth (bw) of h of a given birthdate. Plot A $bw = 14, 15.125$. Plot B $bw = 14$. Plot C $bw = 11.75$. Plot D $bw = 12.14$. Plot D $bw = 14$. Plots A-D Bandwidths chosen as default histogram option. Specifically: number of bins chosen as: $\min(\sqrt{N}, \frac{10 \ln(N)}{\ln(10)})$, where N is number of observations.

Table B.2: Local Linear Estimates of Scholarship on First Year Credits and GPA

Panel A: Dep. Var.: Continuous Credits			
	(1)	(2)	(3)
Scholarship	4.87*	5.61*	0.54
	(2.91)	(3.39)	(2.49)
Corrections	-	B-C, R	B-C, R
Sample	All	All	Ac. Elg.
Bandwidth	134	134	133
Eff. Obs. Left	658	658	393
Eff. Obs. Right	706	706	462
Obs-Left	1970	1970	1081
Obs-Right	1979	1979	1356
Panel B: Dep. Var.: Continuous GPA			
	(1)	(2)	(3)
Scholarship	0.34	0.39*	0.51**
	(0.21)	(0.23)	(0.22)
Corrections	-	B-C, R	B-C, R
Sample	All	All	Ac. Elg.
Bandwidth	193	193	146
Eff. Obs. Left	868	868	407
Eff. Obs. Right	978	978	500
Obs-Left	1865	1865	1055
Obs-Right	1925	1925	1335

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Estimates displayed are the second stage effect of scholarship from fuzzy regression discontinuity design using birthdate in relation to Kindergarten entry date as the cutoff/instrument. The estimate is the effect of receiving the Promise scholarship. In all panels, the first stage cutoff has a strong effect on scholarship receipt. In all results, bandwidths are calculated using \hat{h}_{opt} , in which a common bandwidth is assumed on each side. In columns 1 and 2, the full sample is used. In column 3, only students who exceed the academic cutoffs are included in the estimation sample. In columns 2 and 3, the bias-correction and standard error adjustment in [Calonico et al. \(2017\)](#) are used, because of the additional bias from using MSE-optimal bandwidths. Panel A the dependent variable is a dummy for finishes 30 credits (the renewal requirement) freshman year. Panel B the dependent variable is the number of continuous credits completed freshman year. Panel C the dependent variable is a dummy for exceeds a 2.75 GPA (the renewal requirement) freshman year. In Panel D the dependent variable is continuous GPA.

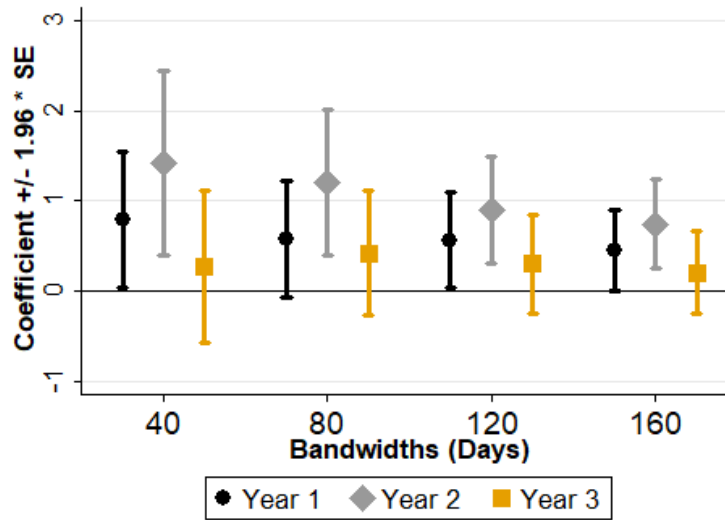
Table B.3: Scholarship on Freshman Credits, Parametric

	(1) +/- 365 Days	(2) +/- 180 Days
Panel A: Credits, Linear		
	(1)	(2)
At Least 1 Year Scholarship	0.17*** (0.03)	0.27*** (0.05)
Observations	2437	1140
Kleibergen-Papp F	4273.3	1646.5
Panel B: Credits, Quadratic		
	(1)	(2)
At Least 1 Year Scholarship	0.21*** (0.04)	0.26*** (0.05)
Observations	2437	1140
Kleibergen-Papp F	2605.5	1924.4
Panel C: GPA, Linear		
	(1)	(2)
At Least 1 Year Scholarship	0.04 (0.03)	0.12** (0.05)
Observations	2437	1140
Kleibergen-Papp F	4273.3	1646.5
Panel D: GPA, Quadratic		
	(1)	(2)
At Least 1 Year Scholarship	0.06* (0.04)	0.12** (0.05)
Observations	2437	1140
Kleibergen-Papp F	2605.5	1924.4

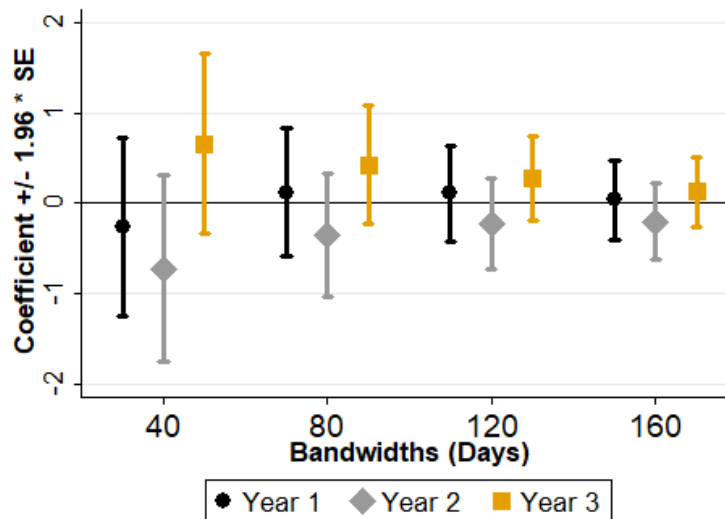
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are robust to heteroskedasticity. Coefficients are from 2SLS regressions, representing the difference at the birthdate cutoff when a functional form for birthdate is assumed globally. In Panel A, exceeding the completed credits (30) in freshman year required for renewal is the dependent variable and linear regressions are fit on either side of the cutoff. In Panel B, the same dependent variable is used as A, but the functional form of birthdate is assumed to be quadratic. In Panel C, the dependent variable is exceeding the freshman GPA (2.75) required to renew the scholarship and linear regressions are fit on either side of the birthdate cutoff. Panel D uses the same dependent variable as Panel C and assumes the functional form of birthdate is quadratic. Finally, column 1 uses 365 days on either side of the cutoffs. Column 2 uses 180 days on either side of the cutoffs.

Figure B.10: GPA Estimates by Year by Gender

(a): Probability of Completing Renewal GPA, Male



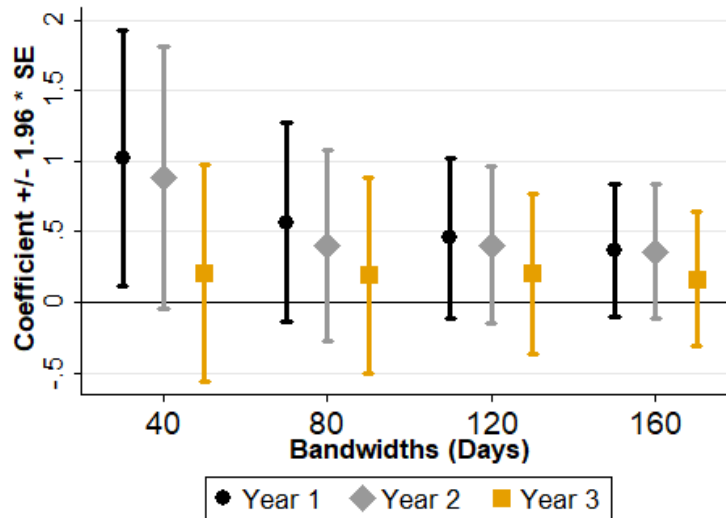
(b): Probability of Completing Renewal GPA, Female



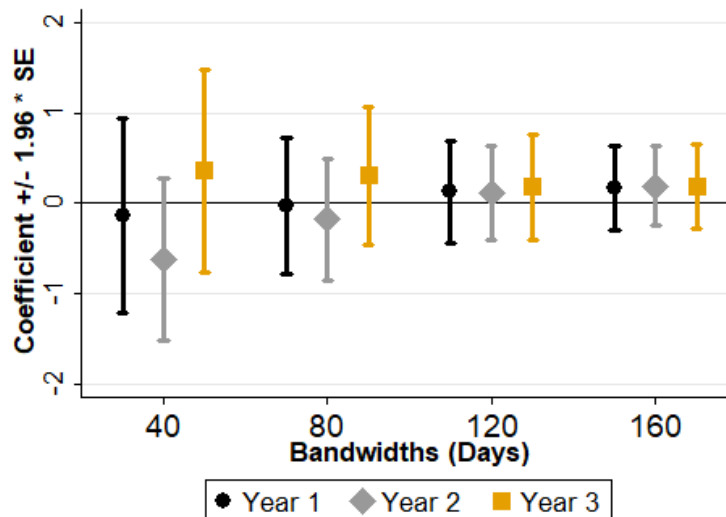
Note: Dependent variable is: finish renewal GPA in that year only. Panel A is male. Panel B is female. $N \leq 524$ in Panel A. $N \leq 541$ in Panel B. Coefficients represent “conventional”, not bias-corrected or robust for MSE-optimal, treatment estimates from difference-in-discontinuity estimates (Grembi et al., 2016). All regressions subsample to academically eligible students meaning [(ACT 20.5 OR SAT ≥ 1000) AND HS GPA 3.0] and include controls for HS GPA and ACT composite score (w/ SAT converted to ACT).

Figure B.11: Credits Estimates by Year by Gender

(a): Probability of Completing Renewal Credits, Male



(b): Probability of Completing Renewal Credits, Female



Note: Dependent variable is: finish renewal credits in that year only. Panel A is male. Panel B is female. $N \leq 524$ in Panel A. $N \leq 541$ in Panel B. Coefficients represent “conventional”, not bias-corrected or robust for MSE-optimal, treatment estimates from difference-in-discontinuity estimates (Grembi et al., 2016). All regressions subsample to academically eligible students meaning [(ACT 20.5 OR SAT ≥ 1000) AND HS GPA 3.0] and include controls for HS GPA and ACT composite score (w/ SAT converted to ACT).

B.2 Compulsory Kindergarten Entry Laws

The compulsory entry laws relevant to this paper are birth date cutoffs for when children must begin kindergarten. These laws vary across states and times. Since Promise begins in 2002 and most students enter college after completing grades K-12, the relevant entry laws were in effect during the 1980's.

There was change in the entry date set by West Virginia's laws over time. In 1982, the rule changed to students who were 5 by September 1 of the current school year were supposed to enter kindergarten. Additionally, in 1988 the law was changed again which means that students in year after are sorted at a different date cutoff, November 1. Thus, the 1988 law cutoff is the last time that September 1 is the cutoff.

The kindergarten date cutoff determining year of college cohort entry depends on how strictly the cutoff is enforced (i.e. whether students born later are allowed to enter early).²⁸ (Whaley, 1985, pg. 20) states, "local districts may permit earlier entry if a child shows readiness based on a test of basic skills." Due to the potential for students to enter early it is highly improbable that the cutoff changes the probability of entering in the first eligible cohort from 0 to 1, thus dictating a fuzzy RD research design. However, (Whaley, 1985, pg. 20) also states, "Complaints occur about the fact one district may allow early entry, while a neighboring one may not," suggesting that the laws are enforced to a certain extent. West Virginia compulsory age of kindergarten entry does not apply to students that enter kindergarten in another state, then move into West Virginia.

B.3 Threats to Identification

There are two main threats to identification. The first is relative age effects, which are addressed by estimating how exceeding the prior year's cutoff, which does not affect scholarship receipt, affects pre-treatment covariates, density, and outcomes. The second is there being another policy that begins with the cohort that begins Kindergarten in 1989.

²⁸If the cutoff is not enforced strictly enough, then the cutoff will not have explanatory power for cohort of entry or scholarship receipt.

Relative Age

An important concern with using a birthdate cutoff in kindergarten entry as variation in scholarship receipt is that the effect of the scholarship is confounded by how a student's position in the age distribution of their cohort also affects outcomes. This effect, of age relative to one's cohort, is called relative age effects and these have been shown to affect many outcomes (Peña, 2017). If one is the oldest in their cohort when they enter kindergarten, then they are the most cognitively developed and this can have a non-trivial difference on outcomes.

Beyond relative age effects, parents have also become aware of this phenomenon and may try to delay the beginning of their child starting school if they are able (D. Deming & Dynarski, 2008). Often knowledge or ability to delay is correlated with income or education which could further confound the effect of the cutoff. Prior work establishes that relative age does matter for probability of admission to college (Matta, Ribas, Sampaio, & Sampaio, 2016) in Brazil.

Relative Age in This Context In this study, relative age could affect the estimated effect of the scholarship in at least two ways. One is a confounding effect where students that are more cognitively advanced when they entered kindergarten have accumulated higher academic ability. Higher academic ability would bias the effect of the scholarship upwards. However, another way relative age could affect outcomes is through selection effects. Students right at the cutoff may be more cognitively advanced and therefore able to attend more prestigious colleges out of state. This would bias the estimates downwards, so the effect of relative age on academic outcomes in this data is ambiguous. While it is possible that students and parents are manipulating their age relative to their cohort, it is implausible that they are purposefully manipulating into college scholarship receipt. Manipulation into scholarship is implausible, because the college scholarship is not announced until 9 years after this kindergarten cohort begins and begins 19 years after the students are born.

Estimating Relative Age Confounding To ensure estimated effects are not due to relative age confounding, this study uses the pre-scholarship cutoff. Students aged 5 on August 31, 1987 are required to enter kindergarten in 1987, but those who turn 5 on September 1, 1987 enter kindergarten in 1988. Thus students born on September 1, 1982 are treated by being the oldest in their class when they begin in 1988. Importantly, these students are not treated by the scholarship. In Section 2.4, students' pre-treatment covariates and the density around the pre-scholarship cutoff (September 1, 1987) will be compared to the post-scholarship cutoff (September 1, 1988).

Other Cohort-Based Policies

A threat to identification is that there is some other confounding policy that begins for children born just beyond the cutoff point (i.e. on or after September 1, 1983). This could be a federal government policy or some other specific policy that affects children in the 1989 kindergarten cohort in West Virginia. To investigate these possibilities, two sources are searched for such programs: 1) the unified welfare analysis across programs in [Hendren and Sprung-Keyser \(2020\)](#) and the record of the West Virginia state legislature.

Policies from a Unified Welfare Analysis ([Hendren & Sprung-Keyser, 2020](#)) There are 133 government policies that are compared in [Hendren and Sprung-Keyser \(2020\)](#). These policies are searched for policies that change discontinuously for students on either side of the birth cutoff, September 1, 1983, for entering Kindergarten in 1989 in West Virginia. While many policies affect students in the 1980's, most do not discontinuously affect those who are born on opposite sides of the cutoff.

Throughout the 1980's there was significant expansions of federal health insurance programs, particularly Medicaid. One policy change was the extension of Medicaid to low-income children who were born after September 30, 1983. This expansion increased education and health care utilization ([Wherry, Miller, Kaestner, & Meyer, 2018](#)); however, this eligibility only changes discontinuously for students who are born 30 days from the cutoff ([Cohodes, Grossman, Kleiner, &](#)

[Lovenheim, 2016](#)). Second, there was an extension of Medicaid to pregnant mothers which raised high school completion for babies who were in utero at the time ([Miller & Wherry, 2019](#)). Nevertheless, while this program changed in the 1980's, it did not affect those on either side of the 9/1/1983 birthdate cutoff differently. This is true for all other health insurance expansions at this time.

C Essay 3

C.1 Additional Tables and Figures

Figure C.1: Survey Taken by NOLS Students

About My Course

Section:

Course Starting Date:
 Month:

Figure C.2: Survey Taken by NOLS Students, Cont.

BEFORE this course Not like me-----Like me	AFTER this course Not like me-----Like me
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
I can assess avalanche slope stability.	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
I can predict ocean tides and currents.	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
I know the land agency's rules and regulations for the region of this NOLS course.	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
I can understand potential solutions to specific environmental problems.	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
I am confident in my knowledge of natural history for the region of this NOLS course.	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
I understand the purpose of Leave No Trace with respect to wilderness travel.	① ② ③ ④ ⑤ ⑥ ⑦ ⑧

For these questions, we only need responses for AFTER.

Disagree-----Agree
① ② ③ ④ ⑤ ⑥ ⑦ ⑧
At least one of my instructors showed a genuine interest in me as a person.
I contributed to my group's successes.
I had important responsibilities on this course.
I made important decisions on this course.
Our group worked well together even when instructors were absent.
I got along well with everyone on this course.
I am very satisfied with my NOLS education.
I am pleased with the pre-course service I got from NOLS admissions.
The NOLS pre-course information described this course well.
Safety was a high priority on this course.

The University of Utah

NOLS

Of these 6 learning objectives, which one did you learn the most about? (Please just choose one.)

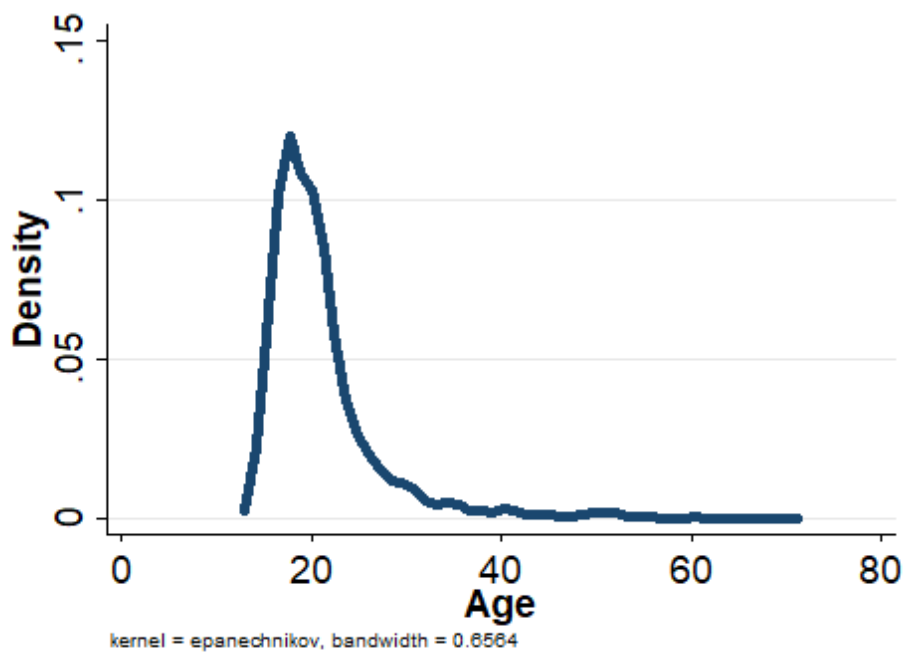
Communication
 Expedition Behavior
 Leadership
 Environmental Ethics
 Risk Management
 Outdoor Skills

Out of all the ways you learned about this objective, which was the most effective and why?

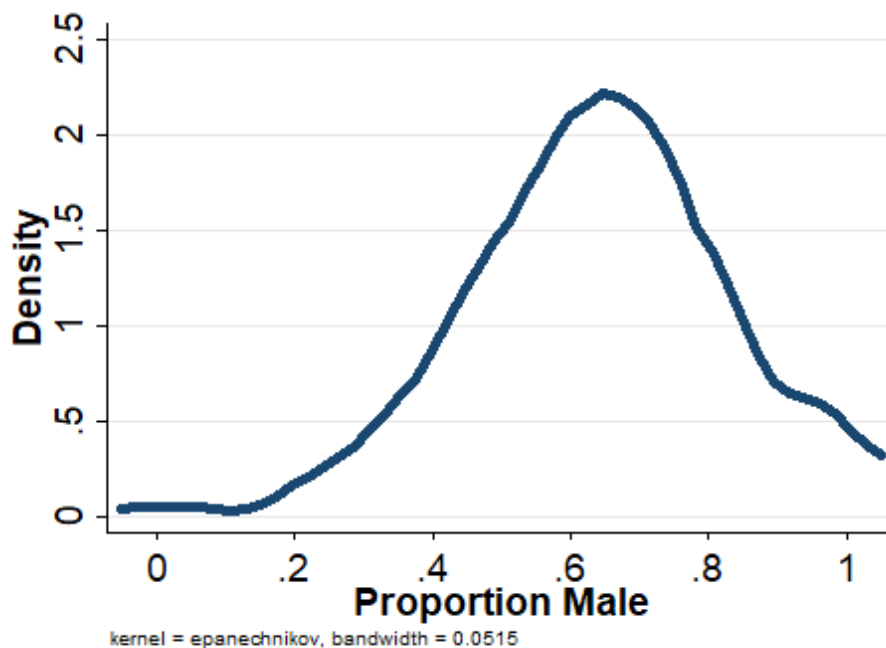
Send this form to:
 U/Utah Research Project
 NOLS 284 Lincoln St.
 Lander, WY 82520.

Figure C.3: Kernel Density of Age and Proportion Male

(a): Age



(b): Proportion Male



Note: Figures show kernel densities. Figure (A) shows kernel density of age, $N = 3,773$. Figure (B) shows the density of the proportion male by section, $N = 358$. There is bump in the right tail of (B), because NOLS makes efforts to ensure that females are not alone in sections.

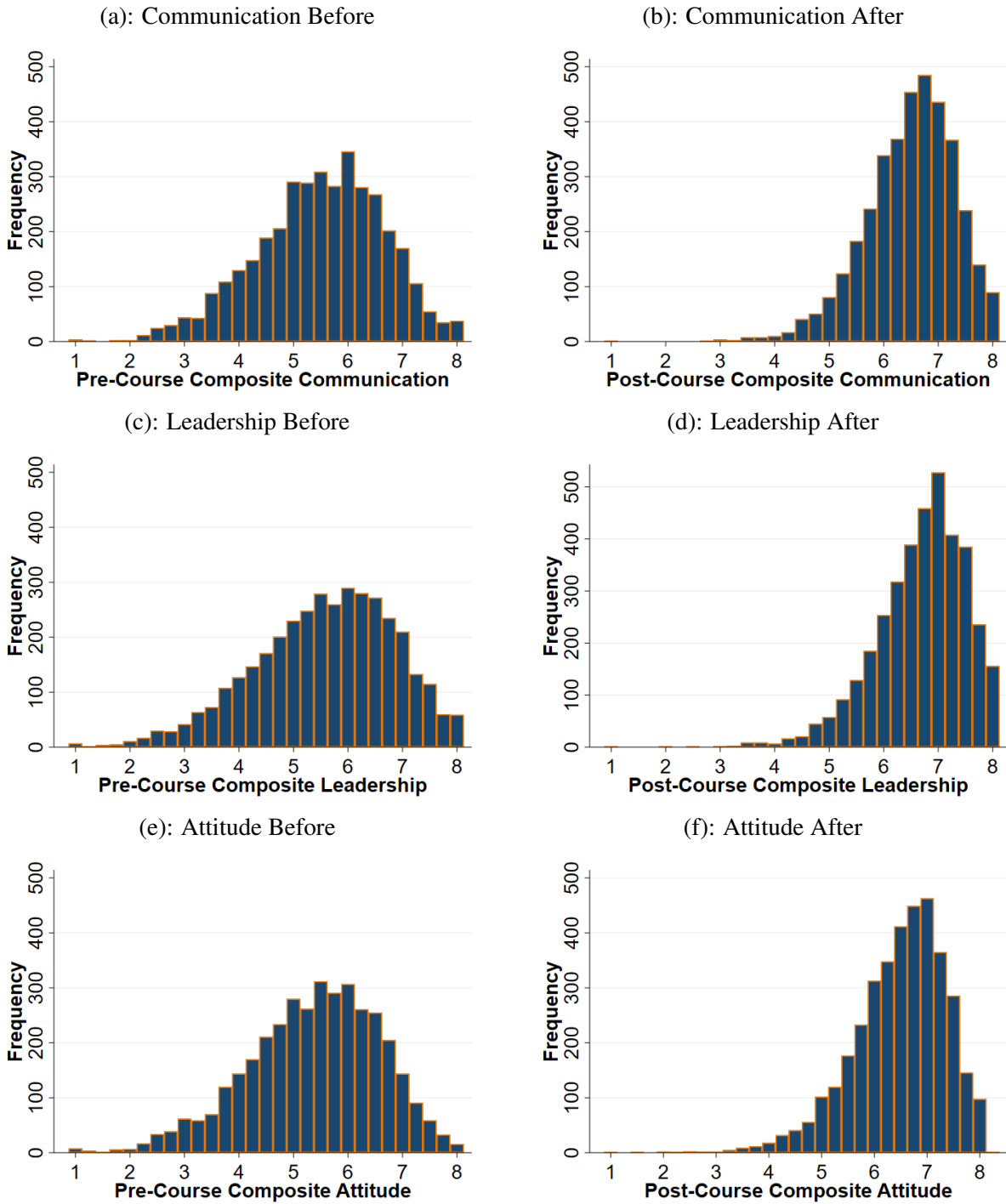
Table C.1: Descriptive Statistics on Raw Questions

Panel A: Before						
	Mean	Median	SD	Min	Max	Count
Communication, Q1	5.44	6.00	1.63	1.0	8.0	3915
Communication, Q2	5.05	5.00	1.62	1.0	8.0	3917
Communication, Q3	5.91	6.00	1.50	1.0	8.0	3918
Communication, Q4	5.52	6.00	1.54	1.0	8.0	3920
Leadership, Q1	5.67	6.00	1.59	1.0	8.0	3918
Leadership, Q2	5.75	6.00	1.62	1.0	8.0	3909
Leadership, Q3	5.55	6.00	1.55	1.0	9.0	3922
Leadership, Q4	5.31	5.00	1.65	1.0	9.0	3910
Attitude, Q1	5.28	5.00	1.77	1.0	8.0	3913
Attitude, Q2	5.09	5.00	1.56	1.0	8.0	3918
Attitude, Q3	5.77	6.00	1.64	1.0	8.0	3919
Attitude, Q4	5.36	5.00	1.50	1.0	8.0	3905

Panel B: After						
	Mean	Median	SD	Min	Max	Count
Communication, Q1	6.54	7.00	1.15	1.0	8.0	3911
Communication, Q2	6.43	7.00	1.15	1.0	8.0	3920
Communication, Q3	6.51	7.00	1.23	1.0	8.0	3916
Communication, Q4	6.46	7.00	1.13	1.0	8.0	3923
Leadership, Q1	6.89	7.00	1.05	1.0	8.0	3919
Leadership, Q2	6.80	7.00	1.07	1.0	8.0	3922
Leadership, Q3	6.64	7.00	1.08	1.0	8.0	3925
Leadership, Q4	6.39	7.00	1.20	1.0	8.0	3922
Attitude, Q1	6.36	7.00	1.36	1.0	8.0	3914
Attitude, Q2	6.23	6.00	1.26	1.0	8.0	3921
Attitude, Q3	6.89	7.00	1.14	1.0	8.0	3919
Attitude, Q4	6.41	7.00	1.17	1.0	9.0	3917

Note: Table shows descriptive statistics for individual survey answers without any aggregation across skills or peer groups. Survey questions are shown in [Table 3.1](#).

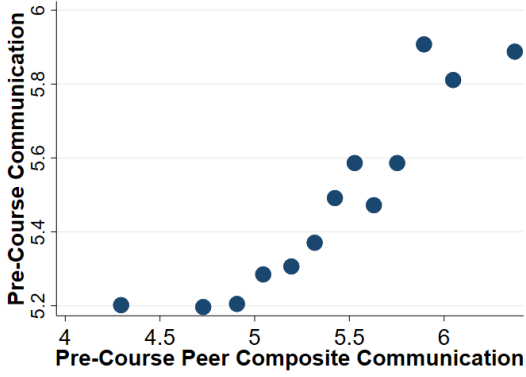
Figure C.4: Distributions of Individual Composite Abilities, Before and After Course



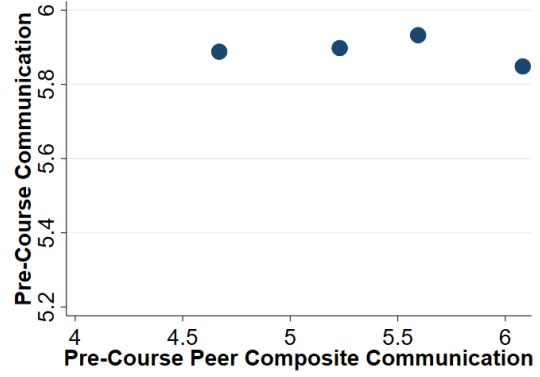
Note: Figures show the distributions of individual composite abilities, before and after the course ends.

Figure C.5: Pre-Course Individual Ability On Pre-Course Peer Average Ability

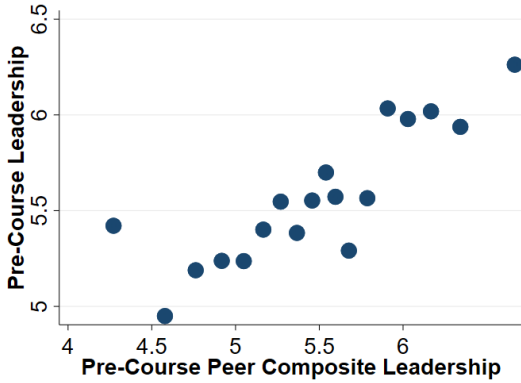
(a): Communication, Unconditional



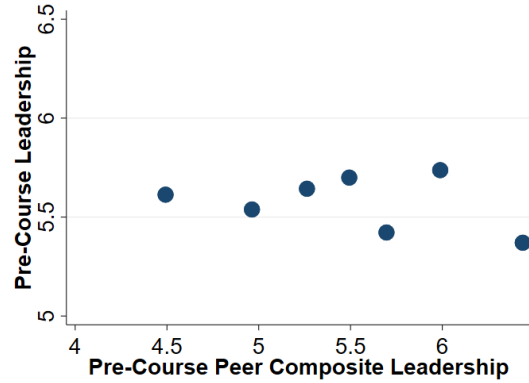
(b): Communication, Course Indicators



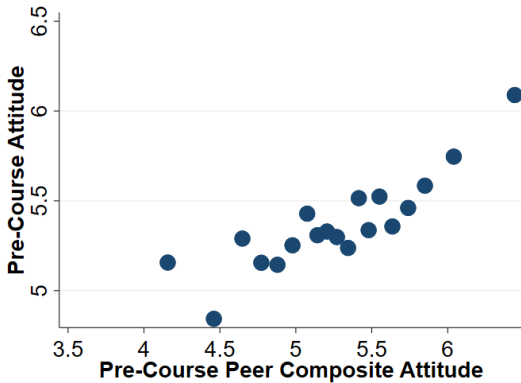
(c): Leadership, Unconditional



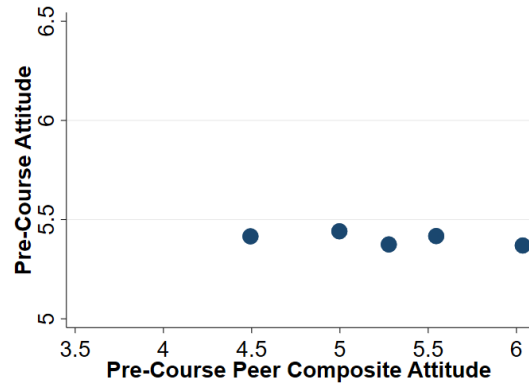
(d): Leadership, Course Indicators



(e): Attitude, Unconditional

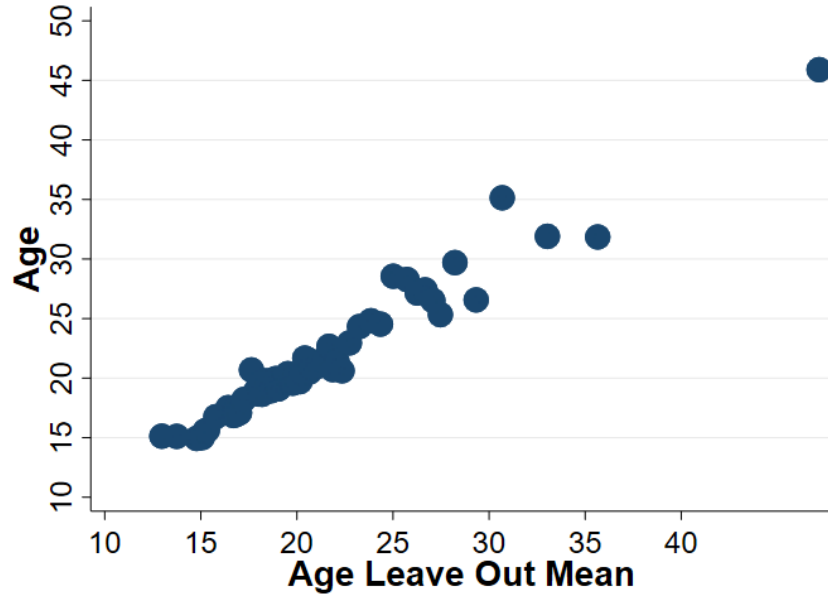


(f): Attitude, Course Indicators

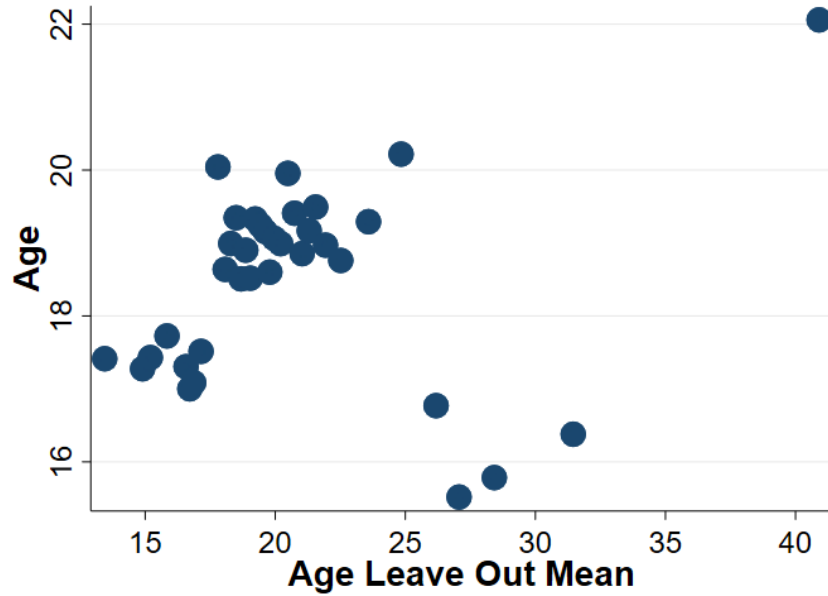


Note: Y-axis is pre-course individual composite ability. X-axis is pre-course peer average composite ability. Number of bins is optimal for variance-bias tradeoff (for integrated mean square error), as defined by Cattaneo, Crump, et al. (2019). Panel A, C, and E are conditional on peer group average and own ability before the course. Panels B, D, and F include course indicators, gender, race, and age as covariates. All figures control for individual pre-course ability. $3,702 \leq N \leq 3709$.

Figure C.6: Binned Scatter Conditional Randomization Checks for Age
(a): Group Age on Age, Unconditional

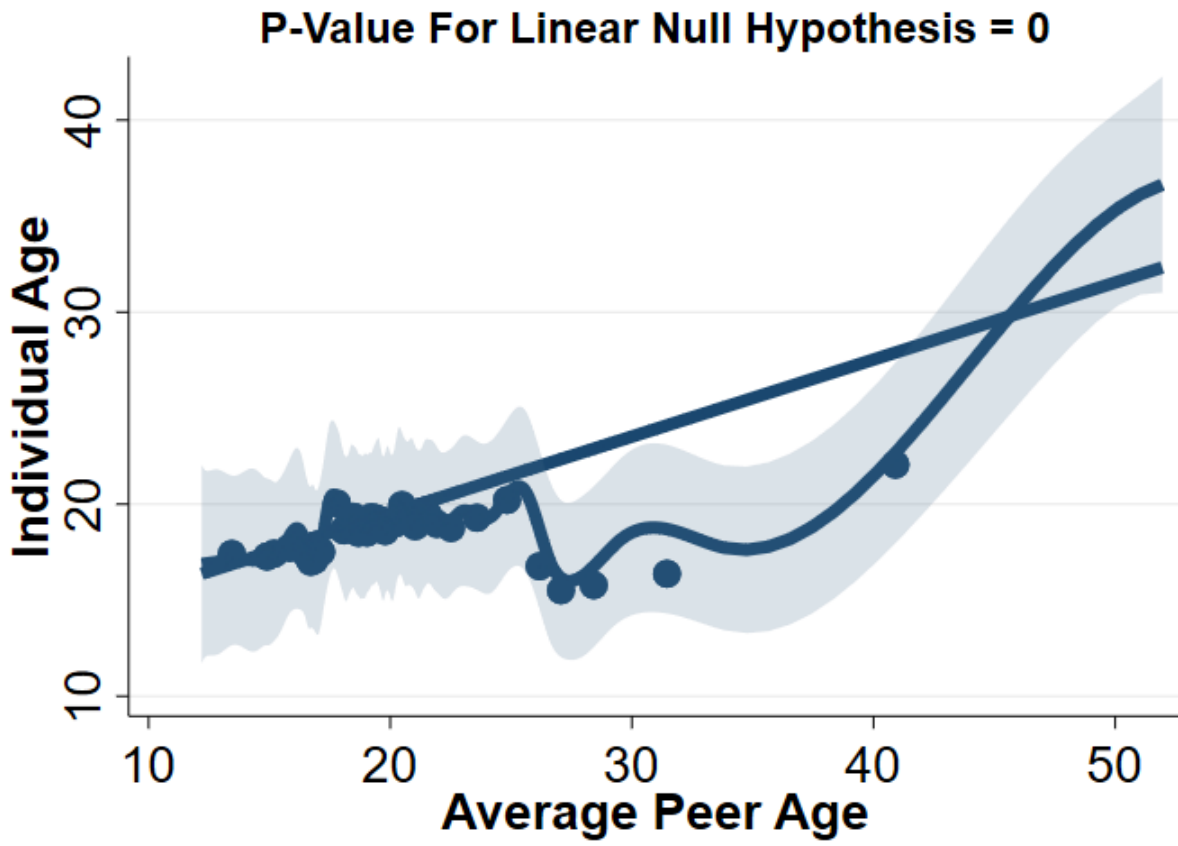


(b): Group Age on Age, Course Indicators



Note: $N = 3,713$ in all figures. Means conditional only on peer group mean. Optimal bins for variance-bias tradeoff (for integrated mean square error) defined by Cattaneo, Crump, et al. (2019). Panel A is unconditional binned scatter. Panels B and C include course indicators as covariates.

Figure C.7: Age on Meanage with Cubic B-Spline



Note: N = 3,773. Fit and confidence bands are from a cubic B-spline. The hypothesis test is a null of global linearity. To evaluate the hypothesis, it compares the linear functional form to a non-parametric functional form, specifically a cubic b-spline.

Table C.2: Relationship Between Peer Average Age and Own Age (Within Mean Age Bins)

	(1)
Age Leave-Out Mean	0.389*** (0.135)
meanage <16 × Age Leave-Out Mean	0.012 (0.061)
16 < meanage < 22 × Age Leave-Out Mean	0.049 (0.045)
Observations	3706
Course Indicators	X

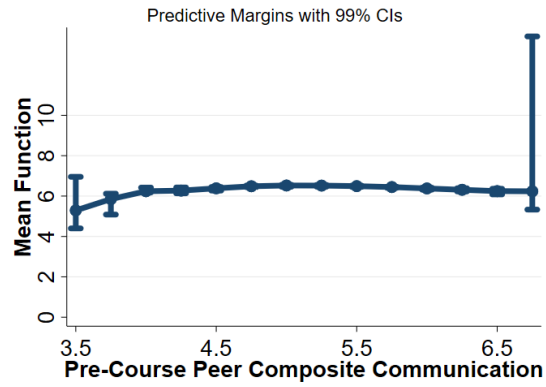
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by courseid, in parentheses. Age Leave-Out Mean is peer average age. Meanage < 16 equals 1 if group average age is below 16, because NOLS has courses that are restricted to only 14-15 year olds. So meanage < 16 X Age Leave-Out Mean is a coefficient for effect of peer age on own age for peer groups below age 16. 16 < meanage < 22 is a dummy variable equalling 1 if the group average age is between 16 and 22, because these are the most common age of NOLS students (see Section 3.1).

Table C.3: OLS of Pre-Course Peer Ability on Noncognitive Ability After, By Individual Question

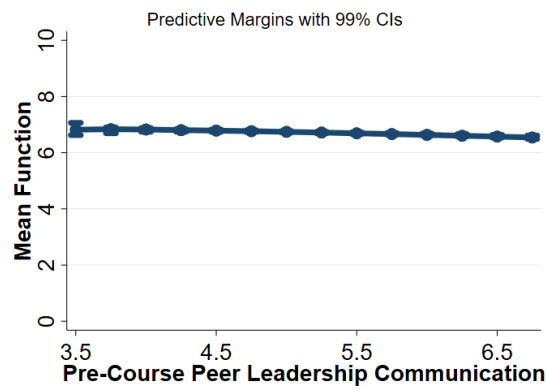
Panel A: Communication								
	Comm 1		Comm 2		Comm 3		Comm 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peer Comm. 1	-0.14*** (0.02)	-0.11*** (0.03)						
Peer Comm. 2			-0.12*** (0.03)	-0.09** (0.04)				
Peer Comm. 3					-0.03 (0.03)	-0.02 (0.03)		
Peer Comm. 4							-0.18*** (0.03)	-0.11*** (0.03)
Observations	3720	3316	3727	3322	3725	3318	3729	3325
R-Sq	0.496	0.524	0.376	0.412	0.615	0.640	0.471	0.501
Individual Covariates		X		X		X		X
Course Indicators		X		X		X		X
Panel B: Leadership								
	Lead 1		Lead 2		Lead 3		Lead 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peer Lead. 1	-0.11*** (0.02)	-0.04 (0.03)						
Peer Lead. 2			-0.08*** (0.02)	-0.07** (0.03)				
Peer Lead. 3					-0.14*** (0.02)	-0.07** (0.03)		
Peer Lead. 4							-0.08*** (0.02)	-0.01 (0.03)
Observations	3726	3321	3728	3322	3735	3329	3723	3323
R-Sq	0.376	0.404	0.422	0.452	0.462	0.492	0.483	0.507
Individual Covariates		X		X		X		X
Course Indicators		X		X		X		X
Panel C: Attitude								
	Att 1		Att 2		Att 3		Att 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peer Att. 1	-0.06** (0.03)	-0.03 (0.03)						
Peer Att. 2			-0.06** (0.03)	-0.00 (0.04)				
Peer Att. 3					-0.07*** (0.03)	-0.04 (0.03)		
Peer Att. 4							-0.06** (0.03)	0.00 (0.03)
Observations	3719	3316	3731	3327	3729	3327	3720	3321
R-Sq	0.454	0.472	0.410	0.434	0.403	0.437	0.422	0.463
Individual Covariates		X		X		X		X
Course Indicators		X		X		X		X

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual communication for a specific question and the independent variable is pre-course peer average communication for the same question. In Panel B, the dependent variable is post-course individual leadership for a single question and the independent variable is pre-course peer average leadership for that same question. In Panel C, the dependent variable is post-course individual attitude for a single question and the independent variable is pre-course peer average attitude for the same question. All columns control for the individual's pre-course ability for the outcome of interest in that panel.

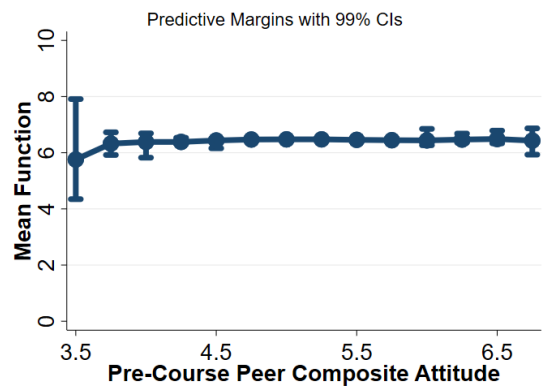
Figure C.8: Marginal Effects From Non-Parametric Kernel Regressions with Confidence Intervals
 (a): Communication Margins



(b): Leadership Margins



(c): Attitude Margins



Note: Figures show the predictive margins of the mean function from the non-parametric kernel regression shown in Equation 3.2 and average effects presented in column 2 of Table C.4.

Table C.4: Non-Parametric Kernel Regression of Peer Group on Noncognitive After

Panel A: Communication		
	(1)	(2)
Pre Comm. Leave-Out Mean	-0.14*** (0.02)	-0.11*** (0.02)
Observations	3658	3265
R-Sq	0.514	0.495
Age Covariate		X
Male & White		X
Course Indicators		X
Panel B: Leadership		
	(1)	(2)
Pre Lead Leave-Out Mean	-0.13*** (0.02)	-0.10*** (0.02)
Observations	3682	3266
R-Sq	0.484	0.566
Age Covariate		X
Male & White		X
Course Indicators		X
Panel C: Attitude		
	(1)	(2)
Pre Att. Leave-Out Mean	-0.08*** (0.02)	-0.04* (0.02)
Observations	3668	3278
R-Sq	0.449	0.429
Age Covariate		X
Male & White		X
Course Indicators		X

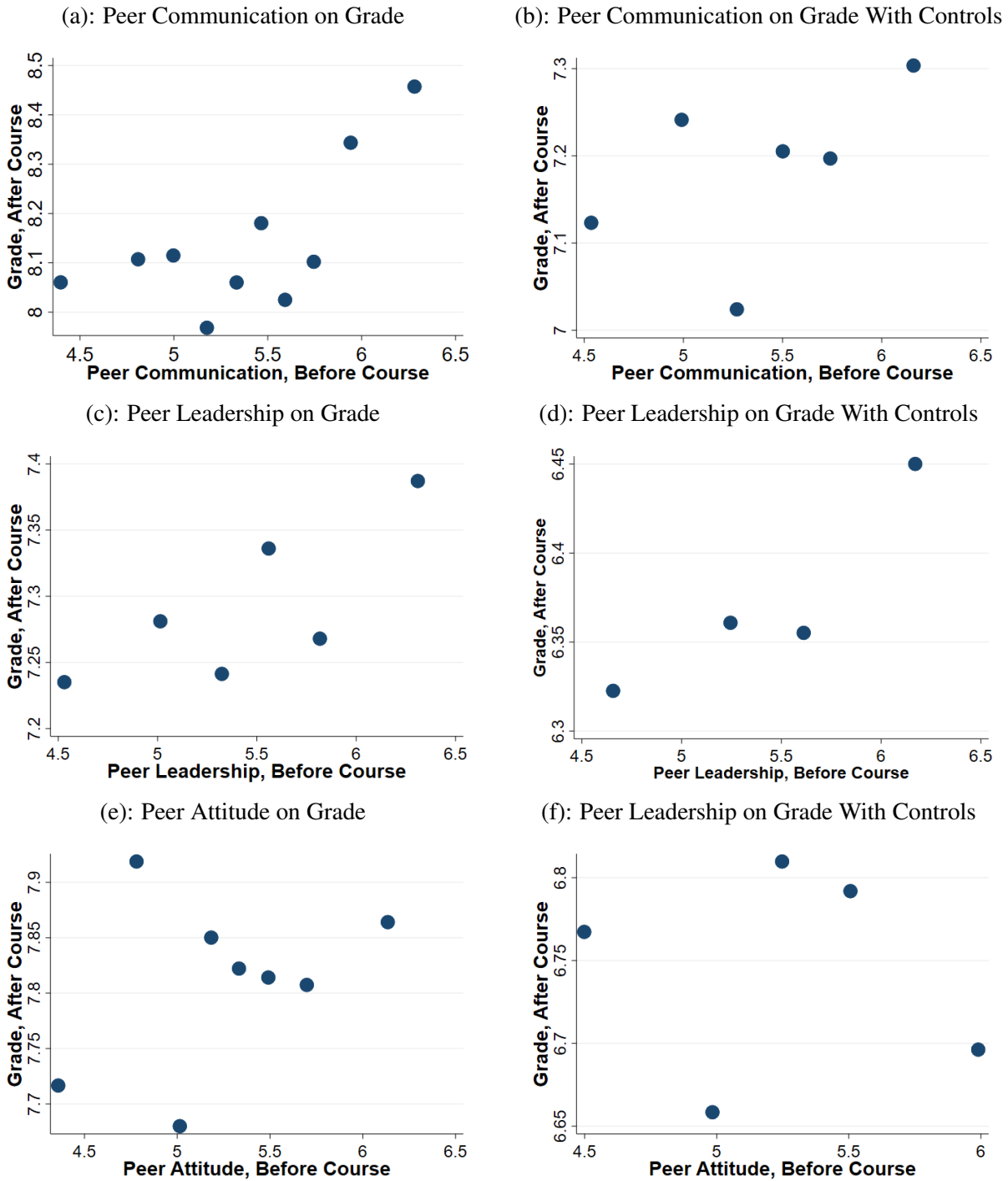
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are bootstrapped for 100 replications. Column 4 uses slightly less replications. All columns include a control for pre-course noncognitive ability. Column 2 adds continuous age. Column 3 adds male and white binary variables. Column 4 includes course type indicators.

Table C.5: Models of Peer Effects

Model	Homogenous Effects	Description
Linear-in-means	Yes	Only the mean of peers background matters
Bad apple	Yes	One disruptive individual harms everyone
Shining light	Yes	One excellent student provides great example
Invidious Comparison	No	Outcomes harmed by presence of better achieving peers
Boutique/ Tracking	No	Students perform best when surrounded by others like themselves
Focus	Yes	Classroom homogeneity is good, regardless of student i 's ability relative to homogenous class
Rainbow	Yes	Classroom heterogeneity good for everyone
Single crossing	No	Positive effects from high ability classmate is weakly monotonically increasing in own ability

Note: Table reproduced from [Sacerdote \(2011\)](#). Models from [Lazear \(2001\)](#) and [Hoxby and Weingarth \(2006\)](#).

Figure C.9: Binned Scatter of Pre Course Peer Ability on Post Course Own Ability, Controlling for Own Before Ability



Note: N = 3,039 in figures A and E. N = 3,035 in Figure C. N = 2,755 in Figure B. N = 2,754 in Figure D. N = 2,759 in Figure E. Optimal bins for variance-bias tradeoff (for integrated mean square error) defined by Cattaneo, Crump, et al. (2019). Panel A, C, and E are conditional on peer group average and own ability before the course. Panels B, D, and F include course indicators as covariates.

Table C.6: OLS of Peer Group on Instructor Assigned Grades

Panel A: Communication		
	(1)	(2)
Pre Comm. Leave-Out Mean	0.15** (0.08)	0.07 (0.09)
Observations	3039	2755
R-Sq	0.0228	0.104
Individual Covariates		X
Course Indicators		X
Panel B: Leadership		
	(1)	(2)
Pre Lead Leave-Out Mean	0.06 (0.06)	0.04 (0.08)
Observations	3035	2754
R-Sq	0.0741	0.153
Individual Covariates		X
Course Indicators		X
Panel C: Attitude		
	(1)	(2)
Pre Att. Leave-Out Mean	0.06 (0.07)	0.06 (0.08)
Observations	3039	2759
R-Sq	0.0354	0.121
Individual Covariates		X
Course Indicators		X
Panel D: Fully Saturated		
	(1)	(2)
Pre Att. Leave-Out Mean	-0.10 (0.10)	0.02 (0.10)
Pre Comm. Leave-Out Mean	0.08 (0.11)	0.03 (0.11)
Pre Lead Leave-Out Mean	0.07 (0.09)	0.02 (0.10)
Observations	2964	2690
R-Sq	0.0794	0.160
Individual Covariates		X
Course Indicators		X

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In all panels, the dependent variable is the recoded integer (1-12) grade. In Panel A, the independent variable is pre-course peer average communication for the same question. In Panel B, the independent variable is pre-course peer average leadership for that same question. In Panel C, the independent variable is pre-course peer average attitude for the same question. All columns control for the individual's pre-course ability for the outcome of interest in that panel. The individual covariates include age, race, and previous experience. The course indicators include course type, section year, and section month.

C.2 Principal Components Analysis

There are 4 questions that are used for each ability. The primary method for combining these questions into information about the underlying ability is to take the average. While an advantage is that it creates a more easily interpretable coefficient for regression analysis, it abstracts away from the potentially different aspects of each ability covered by each individual question.

An alternative method which could be used to reduce the dimensions of each ability is principal components analysis (PCA). In PCA, the components of the correlation matrix are found, such that each component is orthogonal to the others. So an alternative way for combining the various questions into 1 measure of skill by finding these components. While this may not give interpretable regression coefficients, it can at least inform us of whether aggregating the questions, then aggregating by peers is affected by choice of method for aggregating the questions.

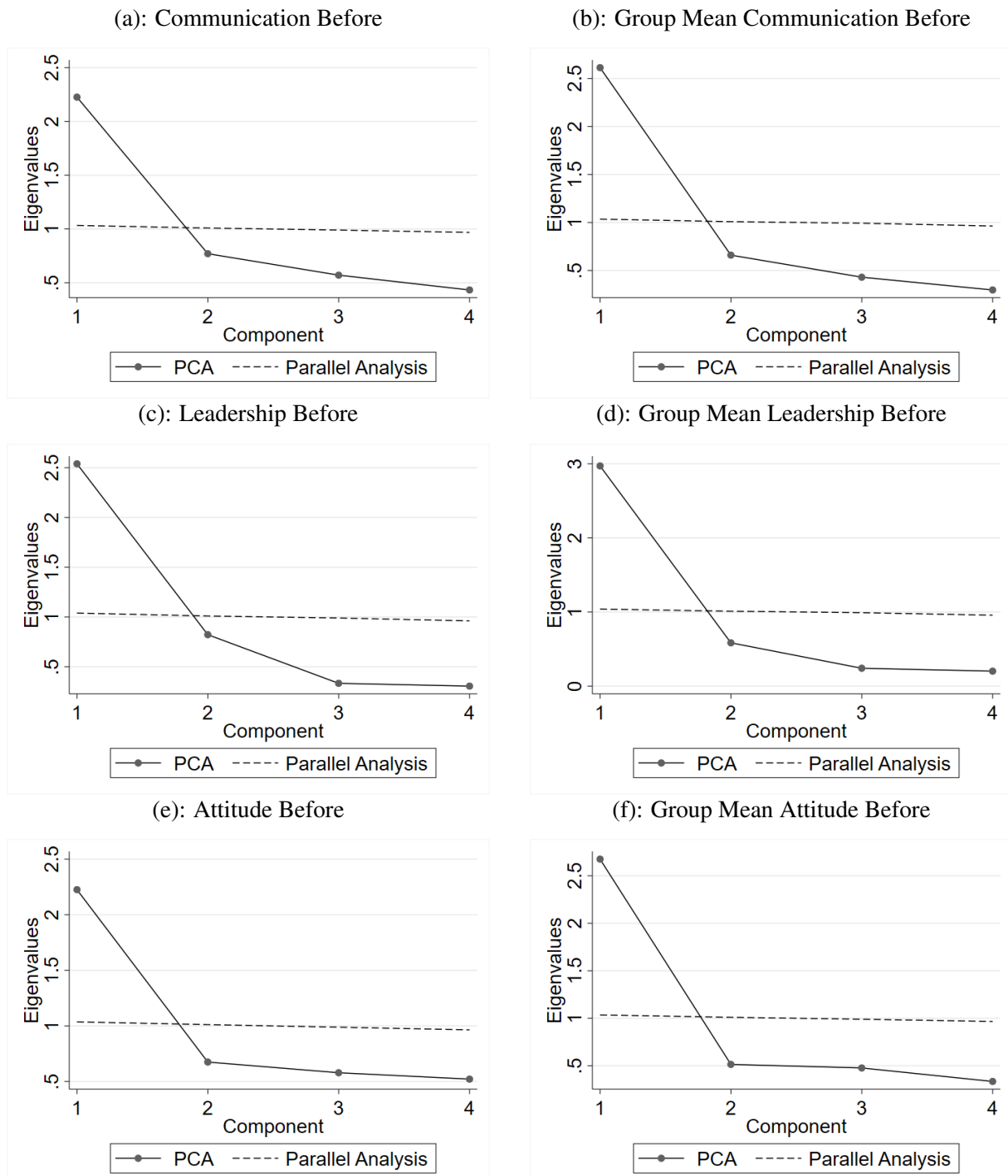
The first step is investigating whether the questions are suitably correlated for a PCA. To investigate this, the Kaiser-Meyer-Olkin (KMO) statistics (which range from 0 to 1) are calculated. Too low of a KMO statistic (typically 0.5 or 0.6) implies too little correlation amongst the variables being considered for PCA. The overall KMO statistics for communication, leadership, and attitude are 0.7384, 0.6904, and 0.7545 respectively, while the individual questions of each are never below 0.686 (lead question 1). The conclusion is that PCA is statistically justifiable as a way to reduce the dimensions of the before individual abilities. Also, the KMO statistics are never below 0.75 once the leave out mean is calculated within groups, making PCA appropriate for the mean group score for each question also.

Second, we are interested in how many components would be required to be retained for regression analysis. [Figure C.10](#) plots the eigenvalues of each component. Sometimes, the so-called “elbow rule” is used to decide how many components to retain, but this can be subjective. Instead, a parallel analysis is performed, which creates a random dataset with the same number of observations and variables as the original data. When the eigenvalues from the random data are larger than the eigenvalues from the pca you know the components are mostly random noise. The parallel analysis crosses the actual eigenvalues plot prior to component 2 in all figures implying that

only 1 component should be retained for regression from either the individual or group means of abilities. [Figure C.11](#) shows that the first component alone always explains about 60% of the data and explains about the same for individual or group mean questions. The first component usually explains a little bit more variance for the group mean than the individual questions.

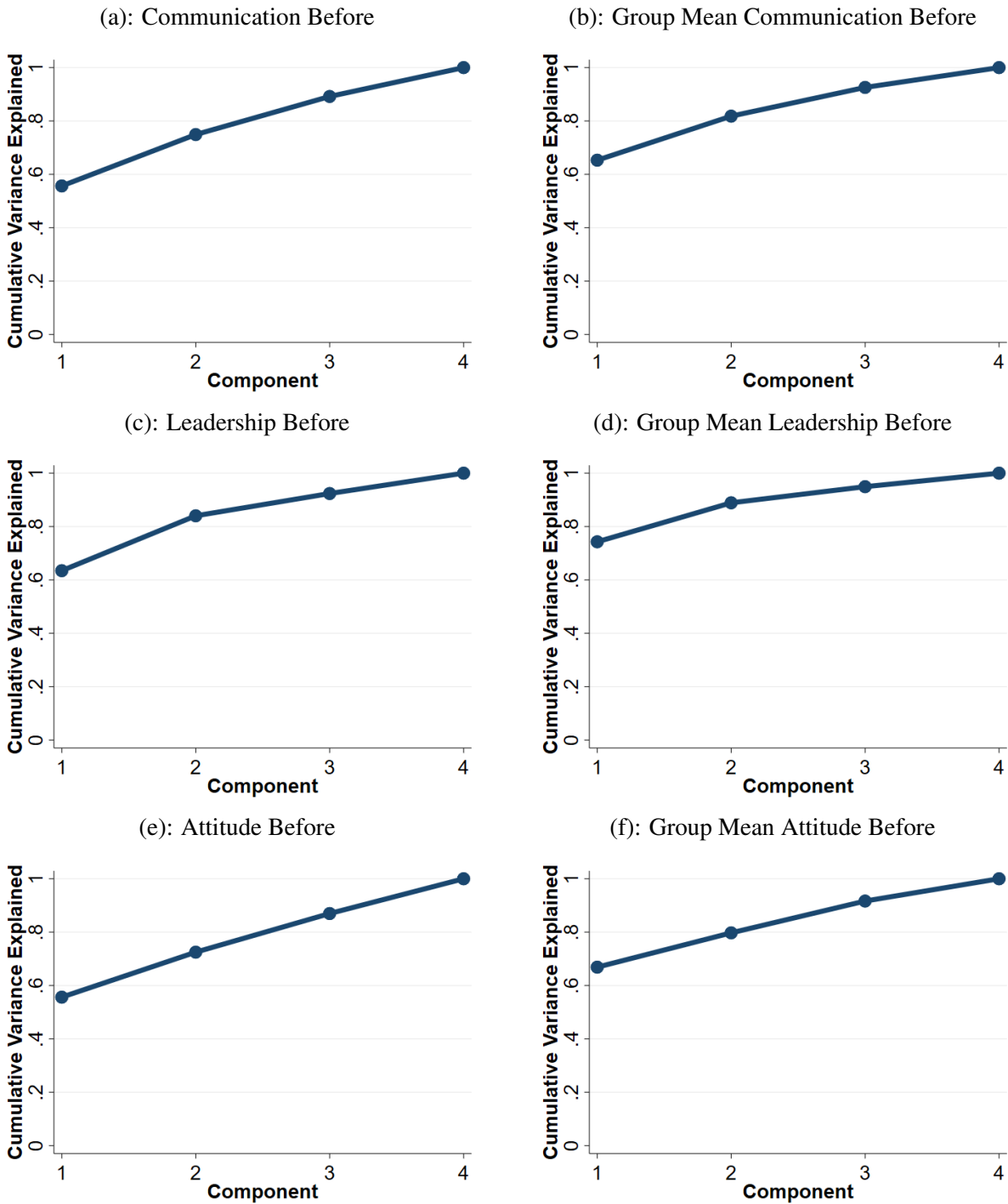
Finally, we can see which survey questions load on which components in [Figure C.12](#). For communication, question 3, which asks about body language, loads heavily on component 2 while all the other questions load heavily on component 1 for both individuals and group means. For leadership, question 4 loads heavily on component 2 and question 2 loads heavily on component 1 in the individual responses. Leadership question 4 loads equally on each component which is the only difference between group and individual leadership questions. Finally for attitude for both individual and group averages, questions 1 and 2 load heavily on component 1 and question 4 loads heavily on component 2. Overall, the analysis concludes that aggregating to group means does not alter how the questions load or how much variance is explained by each component.

Figure C.10: Parallel Analysis



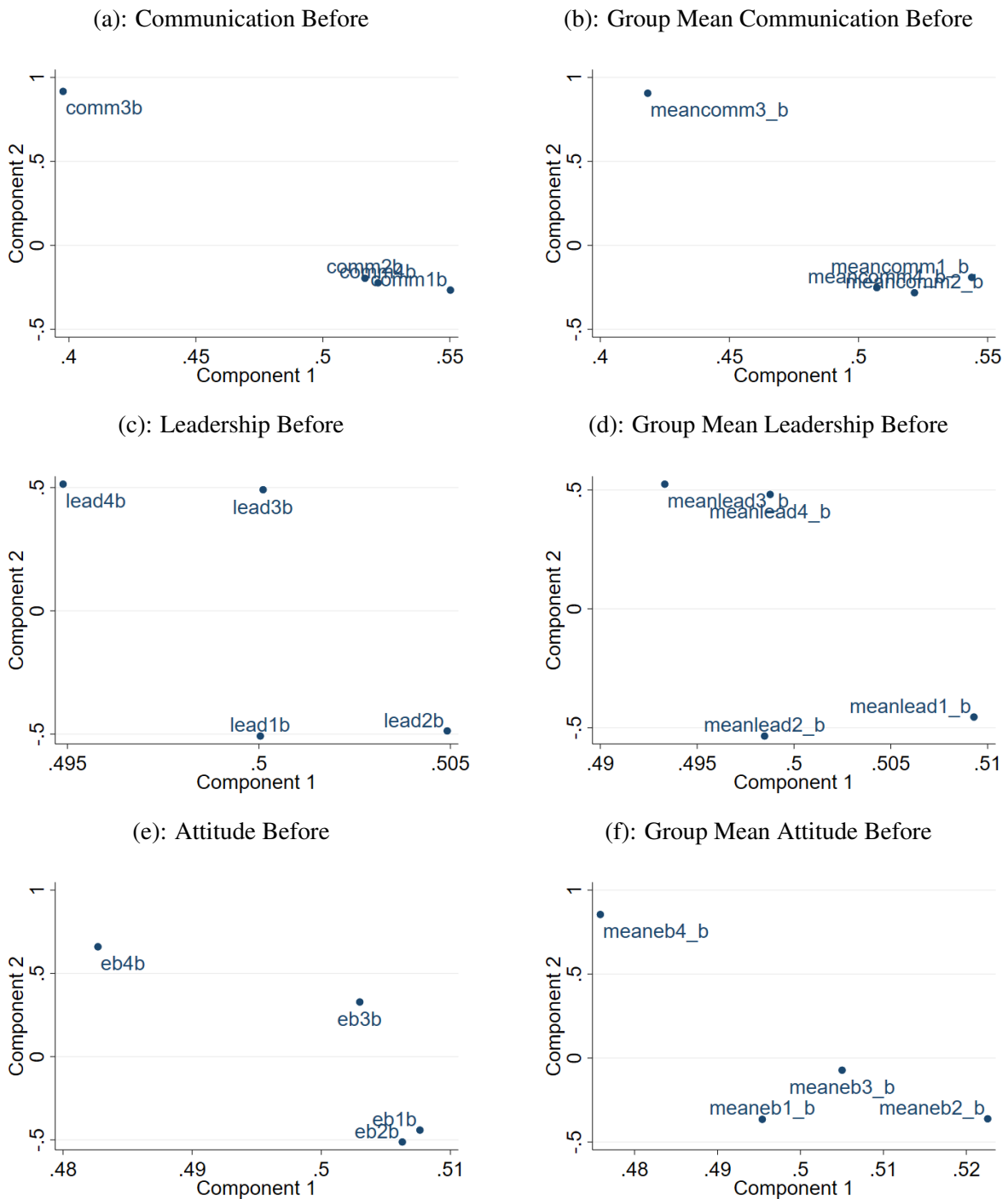
Note: Eigenvalues come from principal components analysis (PCA), of the correlation matrix, for the 4 questions for each ability. The parallel analysis uses 10 repetitions. Figures A, C, and E use the individual questions. Figures B, D, and F use the group means.

Figure C.11: Cumulative Variance



Note: Figures show the cumulative variance explained by each successive component. Panels A, C, and E use the individual measures. Panels B, D, and F use the peer measures.

Figure C.12: Loading on Components 1 and 2



Note: Figures show how each question loads onto each of the first 2 components. Figures A, C, and E use individual measures. Figures B, D, and F use peer measures. Comm3b stands for communication question 3, before the course. Meancomm3_b stands for peer average communication on question 3, before the course.