Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy

TITLE **ICT and Peer Effects on Academic Performance in a University Setting: Evidence from Portugal**

PRESENTED BY **Ryan J. Turner**

ACCEPTED BY THE DEPARTMENT OF

Engineering and Public Policy

Pedro Ferreira — May 5, 2015
ADVISOR, MAJOR PROFESSOR — DATE

Douglas Sicker — May 5, 2015
DEPARTMENT HEAD — DATE

APPROVED BY THE COLLEGE COUNCIL

Vijayakumar Bhagavatula — May 12, 2015
DEAN — DATE

ICT and Peer Effects on Academic Performance in a University Setting: Evidence from Portugal

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Engineering and Public Policy

Ryan J. Turner

B.S., Electrical Engineering & Computer Science, University of Virginia

M.S., Engineering and Public Policy, Carnegie Mellon University

Carnegie Mellon University	Instituto Superior Técnico
Pittsburgh	Lisbon

May 2015

Keywords: Education Policy, Higher Education,
ICT Effects, Peer Influence, Randomization

For Samuel

Abstract

Measuring the impact of educational inputs on academic outputs is a key goal for education policy research. Such measurement can guide the utilization of limited resources and help students be more successful. In this thesis, we study the impact of two important inputs to the educational production function. On the technological side, we study the impact of ICTs, specifically wifi usage and laptop ownership. On the sociological side, we study the effect students have on each other, or the peer effect.

Using data from the Engineering School at the University of Porto, we measure the effect of wifi usage on academic performance using a First Differences model to control for student-specific time-constant effects. We find a positive and significant wifi effect. We find that daytime usage follows overall total effect, while nighttime usage has no effect, suggesting that daytime wifi usage is more academically productive than nighttime wifi usage. We find that the wifi usage in the first curricular year is not academically productive, but that students become increasingly productive as they proceed towards graduation. We also see heterogeneity of the wifi effect among majors, with more ICT-oriented majors seeing the largest positive effect.

We employ a Dynamic Propensity Score Matching model to corroborate the First Differences results, and to include additional non-wifi observations in the estimation of Laptop and total ICT effects. We find positive and significant wifi and ICT effects as before, but a zero laptop effect.

The increasing ubiquity of Information and Communication Technologies allows for measurement at an unprecedented spatio-temporal resolution, and which can be used to position students in relation to each other. In the second part of this thesis, we use session-level wifi data to proxy the social network of students, and employ a randomization strategy to identify a causal effect. We find a positive, statistically significant peer effect that is differentiated by student types. In particular, we find support for a policy of tracking early curricular year students, and grouping later curricular year students.

Acknowledgments

For a dream cometh through the multitude of business;

and a fool's voice is known by multitude of words.

— Ecclesiastes 5:3

First of all, I would like to thank my wife, Ana Patrícia for the amazing friend and companion she is. I am truly happy in her presence and delighted to be married for eternity. I would like to thank my parents for instilling me with the keys to success: the love of learning, and the love of work. I'd also like to thank my in-laws for their tireless support and encouragement.

I extend special appreciation to my advisor, Pedro Ferreira, for inviting me to join the CMU-Portugal community and for guiding the way through the research process. Likewise, I extend deep gratitude to Rodrigo Belo for blazing the trail and for his friendship. In addition, Rodrigo obtained the initial data and wrote most of the initial code to import and process those data, and he frequently contributed other engineering solutions. As is generally the case, this thesis is truly a collaborative effort, and could not be what it is with the extensive support of this research team. In acknowledgement thereto I have chosen to write this thesis in the first-person plural.

Across the pond, I wish to thank my co-advisor at IST, Rui Cunha Marques, whose service and support was essential to the completion of this thesis and for which I am deeply grateful. In addition, I express heartfelt appreciation for the interest, support, contribution, and service of all the members of my thesis committee, Indira Nair and Granger Morgan at CMU and Joana Mendonça at IST.

I express my appreciation to Ricardo Morla at FEUP for providing access to the wifi and administrative data that made this work possible, and to Dossa Massa, for his support in running the queries that were essential for the latter part of this work. I express my appreciation to the Telecom Management and Policy research group seminar for their comments and feedback. I would also like to extend my appreciation to all of the administrative staff in the Engineering and Public Policy Department, most especially Vicki Finney, Adam Loucks, and Barbara Bugosh.

Finally, I thank the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) and the Carnegie Mellon-Portugal Program for funding this research under Grant No. SFRH/BD/51154/2010.

Table of Contents

List of Tables

List of Figures

List of Terms

ATT Average Treatment Effect on the Treated

Bias Difference in covariate means between treatment and control groups

CMU Carnegie Mellon University

FD First Differences

FEUP Faculdade de Engenharia da Universidade do Porto

FCCN Fundação para a Computação Científica Nacional

FE Fixed Effects

ID Identifier (as in Student ID)

GPA Grade Point Average

ICT Information and Communication Technology

OLS Ordinary Least Squares

PSM Propensity Score Matching

RADIUS Remote Authentication Dial In User Service

SNA Social Network Analysis

SAT A standardized test used by most U.S. universities in evaluating students for admission

Chapter 1

Introduction & Literature Review

1.1 Introduction

Many factors combine to affect educational outcomes. Students bring with them innate ability, interest, and desire for work, and are affected by many socioeconomic factors. They enter a school with its own unique history, culture, and access to resources. Measuring the impact of these various educational inputs on academic outputs is a key goal for education policy research. Such measurement can guide the utilization of limited resources, to help students be more successful.

In this thesis we study two important inputs to the educational production function: the technological and sociological effects on student performance. On the technological side, we study the impact of ICTs, specifically wifi usage and laptop ownership. On the sociological side, we study the effect students have on each other; that is, the peer effect.

Improving higher education productivity could have tremendous economic impacts. In this thesis, I focus on two important inputs to the education production function: Information and Communication Technology (ICT) and peer effects, studying this problem in the context of higher education in Portugal. The education production function models an education system as a system with multiple inputs and multiple (but typically fewer) outputs. Our goal is to measure the impact of education as the transformation of inputs (e.g., funding, demographics) on the production of outputs (college placement, lifetime earnings).

ICT is frequently the target of policy seeking to implement reform and spur innovation in education. ICT changes the way students seek information, enhances their ability to search for information, and changes the way students interact with faculty and with each other. indeed make education more productive, but it can also be distracting, and the literature remains conflicted as to what degree ICT is helpful or harmful to education (Ben Youssef & Dahmani, 2008, p. 45). The internet can be both an indispensable resource and a source of distraction—depending upon the choices of the user, but ICT often fails to make the impact its proponents hope for.

Many schools have already or are anxious to deploy Information and Communications Technologies such as wifi networks (Arabasz & Pirani, 2002b), but the literature remains undecided as to whether, and in what contexts, wifi is academically "productive" (Fried, 2008, p. 906). Work must be done to understand the modes of influence by which internet use has effect. Our goal is to better understand the effect of wifi usage, as a new technology, on educational outcomes.

As Information and Communication Technology increasingly makes its mark on education, it opens up new opportunities for both educational intervention as well as measurement. The proliferation of wifi networks increasingly provides spatio-temporal data points on human behaviors, from which student social networks may be inferred. In this work we seek to identify a peer effect between students, and to differentiate this effect by examining the peer effect between the best and worst students. This leads to clear policy implications regarding group assignments, with the goal being to increase total performance or social welfare. Differentiation of heterogeneity will help us to infer which policies would be more effective into improving education.

Finding a null peer effect, Foster (2006) exhorts caution in claiming peer effects in higher education without "substantial empirical or theoretical innovation" (p. 1455). Human factors, like teacher effects, peer effects, and organizational culture, often drive educational outcomes. As in most econometric studies, peer effects are difficult to measure because of endogeneity, insomuch that it is difficult to disentangle the effect of the student on his cohort from the effect of his cohort on the student. Where natural experiments are not available, randomization has been shown to be a promising identification strategy. Randomization allows for the separation of peer effects from homophily and confounding factors, as noted by Anagnostopoulos, Kumar, and Mahdian (2008, p. 8). We believe this robust technique qualifies for the significant technical or theoretical innovation called for by Gigi Foster (2006, p. 1455).

On the output side of the education production function, we simply study assessed grades. We acknowledge the criticism of by some of grades as a measure of academic output, but is accepted herein as an accessible proxy for performance (Allen, 2005).

1.2 Literature Review

1.2.1 Literature Review of ICT Effects in Higher Education

This thesis intersects several large literatures, including higher education policy and technology deployment. A 2011 report warns that "context and fidelity of implementation can matter considerably for the effectiveness of educational technologies" (Council of Economic Advisors, 2011, p. 3). With this in mind it becomes relevant to discuss the specific contexts of internet usage in higher education. We distinguishe secondary education and higher education, wifi and wired internet use, mobile and fixed computing, ICT access at home and school, broadband and narrowband access, and computer-aided instruction and computer-skills training (Angrist & Lavy, 2002, p. 735). These distinctions will help categorize the literature. This thesis focuses specifically on wifi access and on-campus laptop usage in higher education.

Numerous studies have sought to identify the effect of ICT on student performance in various educational contexts. ICT is widely used in higher education, but its impact on student outcomes is not fully understood. We study two particular technologies, wifi and laptops, both of which have seen near universal deployment in higher education, but for which there is little justification (and frequently negative indications) from a program evaluation standpoint. Wifi provides near ubiquitous access to information, while laptops provide a fully mobile platform for intermingling study and sociality. Both technologies are becoming increasingly ubiquitous, and can help students be more productive in a variety of settings.

With these advantages, however, come new opportunities for distraction. In-class laptop usage has been associated with lower levels of understanding and test scores (Fried, 2008, p. 911) and distracting, non-academic behavior (Awwad, Ayesh, & Awwad, 2013, p. 159). Englander, Terregrossa, and Wang (2010) measure the effect of self-reported internet usage on grades and find a negative impact on weekly hours online with actual exam scores in an introductory microeconomics course. In the context of excessive internet use, research finds lower first-year performance (Tindle, 2002, p. 1, Kubey, Lavin, & Barrows, 2001, p. 370), and reduced self-efficacy (Odaci, 2011, p. 1112). These studies are consistent with the zero first-year effect found in this work. These results are tempered by DiNicola (2004), who finds lower rates (10%) of excessive internet usage on a U.S. college campus (p. 99).

Englander et al. (2010) provide an excellent review of literature on the perceived impact of internet usage (p. 86). Empirical research in ICT effects should be preferred above anecdotal evidence and opinions, yet relatively little empirical work demonstrate a clear positive or negative effect in a given context. Since "context and fidelity of implementation can matter considerably for the effectiveness of educational technologies" (Council of Economic Advisors, 2011, p. 3), we provide a brief overview here for both in-class and on-campus dimensions of

ICT usage in higher education.

Much work has been done to survey the perceived effect of self-reported usage. A number of studies examine the effect of ICT usage in one or several sections of a single course. Awwad et al. (2013) conclude that because students use laptops for non-academic purposes that laptops are distracting and should be monitored in class (p. 159). Tindle (2002) concludes that excess internet users underperform academically in their first year of study (p. 1). Both studies are consistent with the zero first-year effect found in this work. Kubey et al. (2001) find a negative effect for discretionary internet usage in the context of internet addiction (p. 370). Odaci (2011) finds that "as problematic [I]nternet use rises, academic self-efficacy declines" (p. 1112).

Some university administrators claim in-class wifi usage is distracting, while others assert that wifi is only distracting when professors fail to engage students (Arabasz & Pirani, 2002b, p. 15, 46). Policy responses to laptop usage are similarly diverse, from mandatory (Olsen, 2001) to prohibited (Chen, 2006).

Fried (2008) discusses the uncertainty we have seen in literature regarding whether or not ICT has a positive effect on student performance, even asserting that there "seem[s] to be a developing feud between those who want to promote laptop use and those who are resistant to it" (p. 906). In a survey of laptop usage, Fried (2008) finds that students who use laptops more during class self-report a lower level of understanding, and also obtain lower grades (p. 911).

For secondary education, Belo, Ferreira, and Telang (2014) examine the effect of broadband internet access on student performance in 9th grade national exams in Portugal. They use a first difference and instrumental variable model, similar to this thesis, to find that a marginal increase in internet usage results in a marginal decrease in student performance. They show that this negative impact is more acute in schools that are lax with respect to how students can use the internet, providing additional evidence that the internet can be quite distracting. This paper is of particular interest due to the shared Portuguese context, and is worth comparison to results in the current work.

Others studies report a very positive effect of broadband internet use, within the construct of well-managed programs. Underwood (2009) has "pointed to well-crafted use of technology benefiting, for example: increased learner effectiveness or performance gains; increased learner efficiency; greater learner engagement or satisfaction; [and] more positive student attitudes to learning (p. 5). The same author in another report finds "significantly improved performances in the year immediately following the installation of broadband"; however, they warn that "[t]his finding should be treated with caution due to small sample size" (Underwood et al., 2005, p. 7). Their overall thesis in both studies is that well-orchestrated programs of broadband have positive educational results: "[i]n short, targeting the use of technology to improving (making more efficient or effective) specific aspects of learning based on a systematic understanding or model will lead to results" (Underwood, 2009, p. 13). They find encouraging results for innovative programs, and is useful as a reference for ICT best practices, but does not adequately consider the general effect of ordinary exposure to broadband internet as a new technology.

In a similar vein, Machin, McNally, and Silva (2007) finds positive effects in British primary schools. They argue that "our positive results [follow from] the joint effect of large increases in ICT funding coupled with a fertile background for making an efficient use of it that led to positive effects of ICT expenditure on educational performance in English primary schools" (p. 2).

Still others indicate the relative unimportance of the technology itself in light of parental and instructor guidance. One study examines the effect of subsidies on home computer adoption in Romania (Malamud & Pop-Eleches, 2011). While the subsidy had a clear effect on the desired ICT result of increased home computer use, it did not increase academic performance. In fact, they conclude that "computers were mainly used to play games," illustrating the importance of purposeful, programmed utilization over mere access to ICT (p. 1024).

We conclude that the effect of wifi as a new, or newly deployed, technology depends strongly on the specific context and framework under which it is deployed and used. Therefore, the work presented in this paper must be understood from the very specific context in which it is conducted. It is hoped that this work will contribute to the overall literature in helping to identify some common elements of "effective" wifi deployments in higher education, according to the goals and objectives of the institution.

1.2.2 Literature Review of Peer Effects in Education

Social networks have been of interest to researchers for many years and to some extent to society for all of humanity. That we influence each other is obvious. But only recently have social networks become digitally encoded through social media and pervasive data-collecting technology. With that comes a trove of insight about the mechanisms by which social influence occurs.

Peer effects in education have received extensive treatment, with both Epple and Romano (2011) and Sacerdote (2011) devoting large effort to methodological and empirical literature reviews. They provide a summary of peer effect models, and discuss several significant econometric issues. They summarize peer effects in education literature, providing a clear and helpful explication of common peer effect models, and discusses identification strategies. He tabulates secondary and higher education peer effects literature.

The canonical work on peer effects in higher education peer effects come from Sacerdote (2001), in which he finds a positive peer effect on GPA at the room level, using a unique data set of randomized roommate assignments at Dartmouth College. Sacerdote provides a useful discussion of endogeneity and identification in social networks.

In contrast, Foster (2006) does not find a peer effect in a unique data set containing both endogenous and exogenous factors and finds the absence of a peer effect at a large U.S. university. She "urges caution in the continued pursuit of peer effects in education without substantial empirical or theoretical innovation" (p. 1455).

The evidence that students affect each other is not surprising. Much more interesting is to discuss which contexts in which they affect each other the most, the magnitude of that effect, and whether it is linear, as there is evidence that peer effects are not linear. Carrell, Fullerton, and West (2009) take Sacerdote (2001) a step further with a data set of exogenously assigned Air Force Academy cohorts, who have very limited social interaction outside that cohort. They find a large and significant peer effect: a 100-point increase in average peer SAT score raises GPA by 0.4 on the standard 4.0 scale. They note that other peer literature has likely underestimated the magnitude of the peer effect since randomly assigned peers constitute a small subset of social interactions. Further, course-level data indicate peer effect is activated through actually working together, and not through the social norm of effort (p. 441). They find persistence of first-year peer effects in subsequent years and nonlinearities based on academic ability of students.

Zimmerman (2003) looks at roommate assignment at Williams College and argues for random assignment in his data set. He finds statistically significant evidence of a causal peer effect, in which students with middle-SAT scores are hurt slightly more by peering with low-SAT score students than they are helped by peering with high-SAT score students.

1.3 Problem Statement

We seek to investigate the effect of ICT and peers on higher education students in Portugal. Given the importance of ICT and peer effects and the difficulty of identifying causal results, this thesis is relevant as a demonstration of the insight that can be gained from these rich data, and in concert with other literature to establish the specific factors that affect student performance in higher education.

1.3.1 Research Questions

This thesis examines ICT effects and peer effects in higher education, and seeks to answer the following questions:

1. What is the effect of ICT usage on student academic performance?
    a) How is the wifi effect differentiated by usage patterns, curricular year, and major?
    b) Is the wifi effect driven by connectivity or mobility?
2. What is the peer effect among students on student academic performance?
    a) How does the peer effect differ across different types of students?

To our knowledge, work presented in this thesis is the first to estimate the actual effect of measured usage with an individualized, campus-wide panel. This work contributes a unique data set and robust, reproducible results. This thesis presents two studies that look at two inputs to the education production function, ICT and peers. Thus, we add to the extensive work that has been done to understand this function. “The key question for most educators is simply whether these technological innovations will have a positive impact on education” (Fried, 2008, p. 907). This research is relevant because literature has not yet come to a consensus on how technology use affects students, and there is some doubt also in the peer effects literature (see Foster, 2006). It is important because our findings can guide policy in a meaningful way; for example, wifi appears to be best employed by more mature students and our results sugest that first-year students could be helped with an ICT resource orientation or by restricting wifi usage in first-year classrooms. Finally, it is interesting because this technology is here to stay and will grow more complex with time, while policy needs to be informed.

1.3.2 About this Thesis

This research has been conducted to facilitate the reproducibility of all results; however, the data are under a non-disclosure agreement and will not be published or made available with this work without consent of the owners of the data. In addition, the anonymity of individual students was preserved at all stages of this research.

This thesis is organized into eight chapters, first addressing ICT and then peer effects. Given the criticality of identifying the specific context of research, we dicuss the context of the thesis in Chapter 2, followed by a chapter on the data. We spend two chapters each on methodology and model, and results, for both ICT and peer effects in Chapters 4-7, and then conclude with policy recommendation in Chapter 8.

Chapter 2

Context of the Thesis

2.1 Overview

The context of this thesis is the Engineering School at the University of Porto (FEUP), in Porto, Portugal. In this chapter, we set the context for ICT deployment at FEUP during five semesters from Fall 2006 to Fall 2008. As will be discussed, this was a time of great transition both for the educational system and ICT deployment at FEUP. The Bologna Process orchestrated changes in higher education curricula throughout Europe so that degrees can be comparable across national boundaries. At FEUP, the five-year undergraduate program was redefined as a 3/2 Integrated Masters, and was implemented between 2006 and 2008. Incidentally, this is also the time period of campus-wide ICT deployment, which had begun as a pilot project in 2001. In this chapter we describe the general features of the secondary school background with which these students entered FEUP, relevant to the study; and, the educational backdrop of the student panels studied in the university.

Education in Portugal is governed by the Ministry of Education, and is much more centralized than in United States. Therefore, on the one hand, it is easier to generalize across all Portuguese institutions, since they are governed by very similar policy. On the other hand, it is more difficult to apply policy implications to institutions outside of Portugal. Nevertheless, we believe meaningful policy implications can be gleaned for application outside Portugal.

FEUP is located in Porto in northern Portugal, and has over 1 million inhabitants in the greater metropolitan area. It is the engineering school in the University of Porto, which consists of the engineering school and several other discipline-specific schools and which serves a total of 22,400 Licentiate and Masters students[1]. FEUP itself has a total enrollment of approximately 7,000 students, which has risen steadily over the period of this study.

2.2 Education in Portugal

2.2.1 Secondary Education in Portugal

While thesis is about higher education, it is helpful to look at the Portuguese secondary school system to better understand the students who go into higher education. There are more than 700,000 secondary students in Portugal. In recent decades, Portugal has seen significant growth in educational attainment, sending 50.2% of its students on to higher education.[2] As in most parts of the world, Portuguese secondary education is geographically determined, although private schooling remains an option.

Broadband was deployed to all secondary schools during time period of this study. In 2001, the Ministry of Education undertook a major policy initiative to connect all Portuguese secondary schools to the internet, a policy they revisited in 2004-2006 to upgrade to broadband internet access. Many of these schools also enjoy open wifi access, although wifi connection policies differ significantly among schools, with some allowing open and unrestricted access and others which set limitations on accessible content and/or schedule.

[1] http://www.estatisticas.gpeari.mctes.pt/archive/doc/insc07_08__difusao_.xls

[2] http://www.nationmaster.com/country-info/stats/Education/Tertiary-enrollment

Internet service is sponsored by the Portuguese National Foundation for Scientific Computing (FCCN). While this initiative no doubt had positive effects in terms of students' ability to access information, and ICT usage skills, Belo et al. (2014) find a negative school-level effect in terms of grades between 2005 and 2009.

2.2.2 Higher Education in Portugal

After the Carnation Revolution of 1974, the Portuguese people began to enjoy new freedoms and greater prosperity. With this came an intellectual revolution in which more and more young people sought and were able to obtain higher education. Portugal, which for decades had struggled with poverty and illiteracy, now sees many of its youth learning to speak English, and obtaining higher education either internationally, or in internationally recognized Portuguese programs.

Colleges and Universities in Portugal

Most Portuguese schools are discipline-specific (e.g., engineering, economics & management, agronomy) and organized under the umbrella of a university. Portugal has among the oldest universities in the world, in particular, the University of Coimbra, which was founded in 1290. Other major university systems include the University of Lisbon and the University of Porto. Public institutions are commonly considered the best institutions, though some of the private institutions join the top ranks, such as Católica University. Top engineering schools include Instituto Superior Técnico, part of the University of Lisbon, and FEUP at the University of Porto.

The process for application and admission to Portuguese universities is standardized throughout the country. Graduating high school seniors take standardized subject-specific exams based on the area of interest, such as Physics or Mathematics in the case of engineering school applicants. Exam scores were averaged together with 11th and 12th grade GPA to yield an Application Score. All students in a given year take the same exams, and so application scores are comparable within any given year, but are not standardized between years. For this reason, we control for cohort and semester fixed effects when controlling for Application Score.

Applicants are given the opportunity to send their scores to six academic programs, sorted by preference. Students are then granted admission at the most-preferred school and program, or major for which they qualify. Obviously, the most highly rated schools and the most popular programs will have the most competition, and therefore will exhibit greater selectivity. Once admitted to a university and major, students have a relatively limited flexibility to change or customize their course of study.

The standard undergraduate degree (first-cycle) awarded in Portugal is called *licenciatura*, or Licentiate, and is awarded after three years of study. Most students continue an additional two years to complete an integrated Masters degree (second-cycle). Doctorates (third-cycle) are not under study in this work.

Bologna Process

In 1999, European education ministers created the European Higher Education Area with an agreement known as the Bologna Process (European Higher Education Area, 1999). The Bologna Process strives to standardize higher education curricula throughout Europe, so that a degree awarded in one country would be comparable to the same degree in another country. Implementation of the Bologna Process was to begin in Portugal in 2006, with mandatory completion by 2008. As such, Portugal saw a phased implementation between 2006-2008. The timing of the adoption of the Bologna Process at FEUP are not published.

The effect of the Bologna Process in Portugal was to convert the standard Licentiate degree from a five-year degree to three-year degree, and to add a two-year integrated Masters degree. Thus, pre-Bologna Licentiate degrees are equivalent to post-Bologna Masters degrees.

As a practical matter, however, implementation of the Bologna Process did not result in a curriculum redesign, as intended, but isntead resulted in the creation of a five-year integrated Masters program. Veiga and Amaral (2009) claim that "the implementation of the Bologna Process in Portugal [has] been achieved in name only" (p. 67). This fact may mitigate the concern of interference in the deployment of ICT.

Performance Evaluation (Grading)

Grades in Portugal, in both secondary and higher education, are given on a 0-20 scale, with grades 9 and below considered failing. The translation to U.S. letter grades may be found in Table 2.1. In Portugal, students are not permitted to drop a course after the first week of class, but it is not necessary to do so because failing grades do not appear on student records. Students generally have limited flexibility in course selection in the first three years, taking a standard set of major courses. There is more flexibility in later years. This is in contrast to U.S. higher education, which typically has a great deal of curriculum flexibility throughout higher education (starting in high school). Students will typically take five courses per semester for the first three years, and four courses or less in the latter two years, in which they focus on writing a thesis.

Table 2.1: CMU-Portugal Grade Conversion

U.S.	Portugal	U.S.	Portugal
A+	19-20	B+	16
A	18	B	15
A-	17	B-	14

ICT Deployment in Higher Education

Wifi deployment at FEUP began as early as 2001 as a pilot project with limited usage Information is scarce as to how this pilot project evolved into a campus-wide network. Logging of wifi usage began in Fall 2006, and the number of users has increased steadily since. Wifi is deployed throughout the small, urban campus via 205 Access Points, which are located in classroom buildings, student study areas, cafés, and other research and study areas of campus. The wifi network at FEUP is part of the Eduroam network, which provides roaming internet access at participating institutions across Europe, and increasingly around the world. Eduroam relies on a distributed authorization protocol called Remote Authentication Dial In User Service (RADIUS) protocol, which also provides the network accounting data used in this thesis. Unfortunately, these data only contain high-level usage information such as session duration and megabytes transferred.

2.3 Comparison with U.S. Education

In addition to the differences already noted, it is worth emphasis that the Portuguese education system is very different from the U.S. system in a number of ways in both secondary and higher education. From a socioeconomic standpoint, Portugal is a relatively more homogeneous society, though immigration into Portugal has risen considerably since decolonization in the mid-1970s. Education policy in Portugal is equally homogeneous through the country, since it is administered nationally, whereas in United States it differs significantly from state to state and across universities. Portuguese students tend to have a more set curriculum, and cannot drop courses for poor performance. In the U.S., students often attend universities in a different city, and live on campus at least for the first year. Portuguese students will also move to a city for university if necessary, but they are more likely to attend university in the same or nearest city. Portuguese students typically live at home and commute, and as a result they spend relatively less time on campus. In the U.S., student information such as grades are strictly confidential, whereas in Portugal, class grades are publicly distributed.

In the U.S., students are generally graded relatively, or on a curve, whereas in Portugal grades are generally absolute. Curriculum is usually more flexible in the U.S. and fixed in Portugal, at both the secondary and tertiary levels. Portuguese teachers are viewed as essential public servants, but are not awarded special status in society, as in Finland. At the university level, a major goal of the institution is research. Portugal in particular has in recent years sought to develop its research program with greater international exposure through partnerships with several U.S. institutions. This thesis is the product of one such collaboration.

While these analyses are of Portuguese students, we believe that much of our results are generalizable to the U.S., notwithstanding these significant differences in education policy. This is because we are investigating inputs to the education production function that are, for the most part, the same in any part of the world. While there are differences for in how and where students get online, Portuguese and American students are increasingly part of the same, English-speaking online culture, and we have no reason to believe that the differences cited would produce substantively different results in the U.S. Some added caution is warranted in applying the peer effects work to a U.S. context because of the wider sociological differences that exist. For example, Portuguese students are more likely to know each other from high school than U.S. students (which only rarely know other individuals on a new university campus), and the reported peer effect might thereby be misattributed. Nevertheless, in general terms we think that both parts of this thesis would be informative to a U.S. higher education policy reader.

Chapter 3

Data

3.1 Overview

This thesis is based on data from the Engineering School at the University of Porto (FEUP) in Porto, Portugal. In this chapter, we outline the data sources, assumptions, and processing, that lead to the data sets under analysis, and provide informative descriptive statistics. Some additional data are generated to support analysis; however, discussion of these methodologies will be deferred to Chapters 4 and 6.

Throughout this thesis, it is important to recognize that the data cover the period 2006–2008, which is coincident with the implementation of the Bologna Process in Portugal, as discussed in Section 2.2.2. The period of the study is a time of transition—of the programs themselves, and the initial campus-wide deployment of wifi. This study, then, occurs during a very dynamic period in terms of both curriculum and ICT deployment.

3.2 Methodology

All data for this work come from FEUP, and are composed of three data sets: student administrative data, course grades, and student wifi usage. The former two are extracted from the student information system at FEUP. Wifi data are provided by the Eduroam network, a network that permits roaming across wifi networks at participating institutions of higher education. These data are processed, aggregated, and combined into two separate data sets: a student-semester panel, and a social network which captures student relationships.

Table 3.1: Description of Variables

Name	Description
Student ID	Anonymized student identifier
Semester	Index by semester from Fall 1999
Grade Points	Cumulative grade points earned per semester
No. Courses	Number of courses completed per semester
Total Hours	Total Hours Online per Semester
Total Megabytes	Total Megabytes Transferred per Semester
Hours (Day)	Daytime (8a-8p) Hours Online per Semester
Megabytes (Day)	Daytime Megabytes Transferred per Semester
Hours (Night)	Nighttime (8p-8a) Hours Online per Semester
Megabytes (Night)	Nighttime Megabytes Transferred per Semester
Application Score	Score used to evaluate student applications for admission
Cohort	Year student entered FEUP
Major Dummies	Indicator variables for each of five engineering majors
Curricular Year Dummies	Indicator variables for curricular year
Semester Dummies	Indicator variables for semester

Academic performance is measured by grades. In addition, we want to measure the total amount of academic work accomplished, and not simply the average performance on work attempted. For this reason, we measure academic performance in terms of number of courses completed, No. Courses, and semester grades points, defined as: Grade Points = Semester GPA $\times$ No. Courses. This allows a better ranking of how much academic work a student accomplished in a given semester. A measure of pre-university performance is given by Application Score, which ranks the quality of students when they enter FEUP, and is the primary method of ranking and admitting students for all Portuguese universities. Table 3.1 shows relevant variables from the merged data. Figure 3.2 summarizes student-level descriptive statistics.

Table 3.2: Student Summary Statistics

	mean	median	sd	min	max	count
Student ID	–	–	–	–	–	17881
No. Courses	3.916	4	2.218	1	35	17629
Grade Points	51.26	50	30.71	10	646	17629
Hours	75.09	37.05	97.80	0.00389	1055.9	5448
Hours (Day)	63.92	32.16	82.74	0	703.7	5448
Hours (Night)	11.18	1.936	27.10	0	750.2	5448
Megabytes	6099.4	1326.8	15173.5	0.000612	289384.0	5448
Megabytes (Day)	5162.1	1122.4	13072.6	0	266163.2	5448
Megabytes (Night)	937.3	33.15	3952.8	0	118871.3	5448
Application Score	149.9	150.8	19.85	15	200	9182
Cohort	2003.7	2004	2.906	1999	2008	12524
EE Major	0.197	0	0.398	0	1	17881
CS Major	0.0794	0	0.270	0	1	17881
CHE Major	0.0649	0	0.246	0	1	17881
ME Major	0.129	0	0.335	0	1	17881
CE Major	0.189	0	0.392	0	1	17881
Other Major	0.341	0	0.474	0	1	17881

Outliers are treated pragmatically. Physically impossible cases (such as having more hours of usage than time in a semester) are removed, but most other data remain intact (including students reporting an impossible number of credits earned in a semester, which may be the result of true accounting such as the application of transfer credit). All data are linked by student ID prior to anonymization. Key variables are standardized to simplify interpretation of results.

3.2.1 Grade Truncation

Grade Points is defined as sum of *passing* course grades in each semester. It follows that some failing grade points earned are not reported—that is, Course Grades are truncated below 10. Following the discussion in Chapter 2, failing grades are somewhat more common in Portugal than in the United States, because students cannot "drop" a course in which they are performing poorly or which they do not intend to complete. The corollary is that failure carries less stigma, and ultimately is not reported in final records.

Unfortunately, this creates an downward bias in measured Grade Points (and an upward bias in measured GPA). That is, if true and measured Grade Points are denoted Y^* and Y, respectively, then $Y < Y^*$. This potentially leads to overestimation of the ICT and peer effect, providing merely an upper bound when there may in fact be none.

We address this problem by defining a new measure, denoted Grade Points*, or Corrected Grade Points, as follows:

$$\text{Grade Points}^* = \begin{cases} \text{Grade Points} + 9 * \max(5 - \text{No. Courses}, 0), & \text{if Curr. Year} \leq 3 \\ \text{Grade Points} + 9 * \max(4 - \text{No. Courses}, 0), & \text{if Curr. Year} > 3 \end{cases}$$

This formula adds 9 points—the maximum failing grade—for each course not passed in an expected load of 5 courses for Licentiate students and 4 courses for Masters students. No correction is made for those students who pass more than the expected number of courses. Thus, Grade Points* by definition is an upper bound measure of the true Grade Points, and yields a lower bound estimation of the ICT and peer effect. We examine all results in terms of both Grade Points and Grade Points* (and typically do not find a statistically significant difference).

3.2.2 Generation of Student Panel

We first aggregate by student over semesters and show student-level summary statistics in Table 3.2. The panel index is the *student-semester*, yielding one record per student for each semester in which that student took courses. Statistics for student-semester will follow. We construct a student-semester panel of grades, wifi usage, and administrative covariates (such as Application Score and Cohort dummies) from separate data sources. Course grades and administrative data are available from 1999 through 2010 (22 semesters), yielding 17,881 students, though not all covariates are available for all observations. Figure 3.1 shows how missing variables overlap to reduce the total number of observations.

The Eduroam network is based on the RADIUS-powered roaming authentication service, which also provides accounting logs reporting session-level connection time and bytes sent

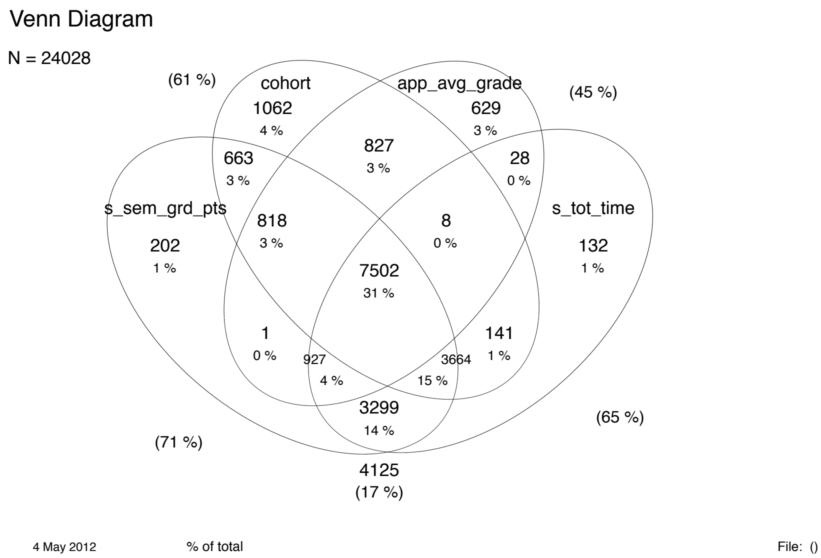
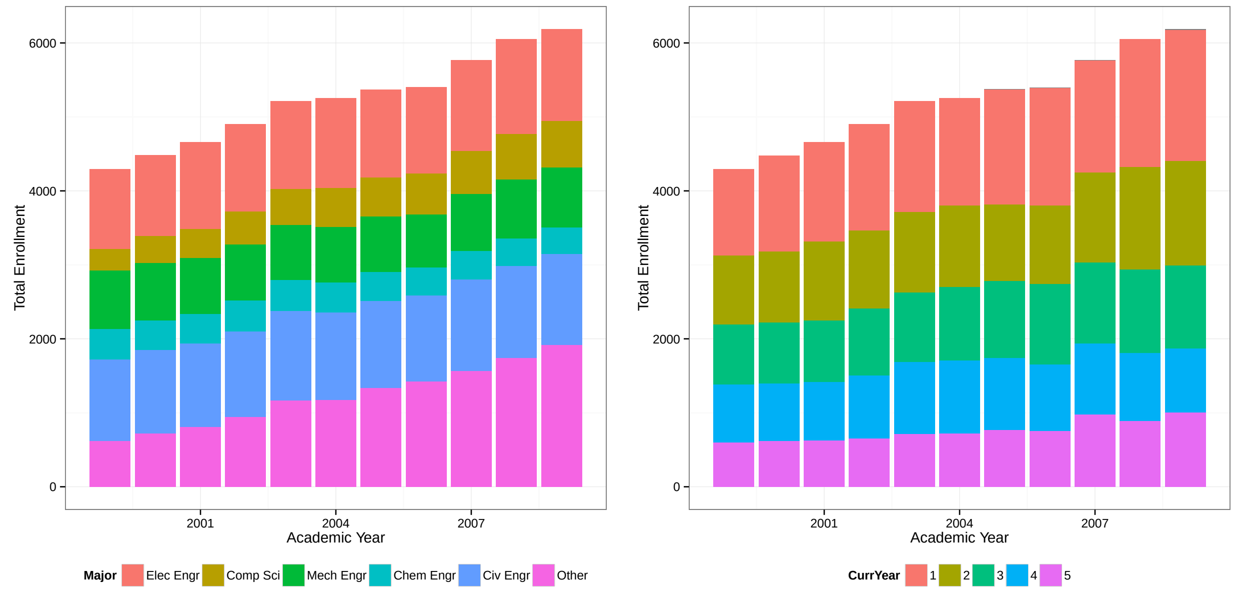


Figure 3.1: Covariate Coverage across Students

and received (see Rigney, Willens, Rubens, & Simpson, 2000 and Rigney, 2000 for more information), but not the nature of content accessed. Ethernet usage data are not available; however, Arabasz and Pirani (2002a) report that "students often choose wireless network access even when wired network ports are available" (p. 9). A 2012 survey indicates that of the 85% of students who use a laptop on campus, 97% of them access the internet via wifi (compared with 8% who access via Ethernet). Wifi was first installed at FEUP in 2001, but reporting did not commence until Fall 2006. Incomplete wifi data from Spring 2009 are dropped, yielding five complete semesters of wifi usage. The data are reduced to 6,425 wifi-using students for the five complete semesters of wifi.

3.3 Descriptive Statistics

Figure 3.2 shows the number of students admitted each year by majors and curricular year, respectively. Student-level descriptive statistics were shown in Table 3.2. Table 3.3 gives First Differences for a relevant subset of panel variables. Table 3.4 provides a correlation



(a) Major (b) Curr. Year

Figure 3.2: Annual Enrollment: The total number of students enrolled in each year, segmented by Major and Curr. Year

matrix for first-differenced variables, as variables appear in the models.

Table 3.3: Panel Summary Statistics—First Differences

	mean	median	sd	min	max	count
Student ID	–	–	–	–	–	6425
Semester	17.65	18	1.091	16	19	6425
ΔNo. Courses	-0.256	0	1.897	-21	6	6425
ΔGrade Points	-2.676	0	25.65	-247	82	6425
ΔHours	27.91	10.87	122.2	-857.7	1178.0	6425
ΔMegabytes	2339.2	268.2	22708.9	-293963.6	279968.7	6425
Application Score	149.1	148.8	17.04	97.50	199	5410
Cohort	2003.9	2004	1.984	1999	2007	6034
Curr. Year 2	0.196	0	0.397	0	1	6425
Curr. Year 3	0.231	0	0.421	0	1	6425
Curr. Year 4	0.254	0	0.435	0	1	6425
Curr. Year 5	0.216	0	0.411	0	1	6425
EE Major	0.287	0	0.453	0	1	6425
CS Major	0.170	0	0.376	0	1	6425
CHE Major	0.0705	0	0.256	0	1	6425
ME Major	0.149	0	0.356	0	1	6425
CE Major	0.165	0	0.371	0	1	6425
Other Major	0.158	0	0.364	0	1	6425

Semester is indexed from Fall 1999, so Fall 2006 = 15. Semester 15 is subsumed in taking first differences.

3.3.1 Student Panel

Performance is measured in terms of Grade Points earned and Number of Courses (No. Courses) passed each semester. These provide an absolute measure of academic work accomplished, accounting for both credit quantity and quality. In Portugal, course grades are given on a 0-20 scale. Grade Points has some censoring bias since failing grades (< 10) are not reported, which is addressed in part by looking at the number of courses (No. Courses) successfully completed each semester. Students generally attempt 4-5 classes per semester, and pass 3.9 classes on average. They earn $4 \times 13 = 52$ grade points per semester, obtaining an average score of 13 (C+). Students in Portugal cannot drop classes after the start of the semester, and consequently have a higher failure rate. This makes the number of passed courses (No. Courses) a much more meaningful measure.

Wifi usage is aggregated by semester in units of Hours online and Megabytes transferred. Students use on average 55 minutes of wifi per weekday during a 157 day semester, transferring about 80 megabytes of data. This is equivalent to approximately 20 minutes of standard quality YouTube video, 130 page views, or 800 emails. Portuguese students rarely live on campus (as on-campus housing is not provided), and can be expected to have low rates of ICT utilization. Note that wifi usage distributions follow a power law relationship, so the median is significantly less than the mean and the standard deviation includes zero.

Application Score consists of the weighted average of 12th grade GPA and scores on national subject exams, such as Physics and Math, and is not standardized. It is used to decide university admission in Portugal, and is comparable to the Scholastic Assessment Test (SAT). We control for semester and cohort fixed effects, in part because scores are not directly comparable across cohorts. They are likely correlated with important unobserved covariates, like income and parent's education, and is used to proxy for expected student performance at university (Rothstein, 2004).

The majority of students in this time frame complete five-year integrated Masters degrees as shown by the five Curricular Year dummies in Table 3.3. FEUP has five standard engineering majors and an "Other" category for smaller engineering majors, with Electrical Engineering (EE) being the most popular major.

3.3.2 Inference of Student Graph from Wifi Data

We measure the magnitude of peer influence using co-located wifi sessions as a proxy for the social network among students. We use this network and randomization to infer the magnitude of peer influence among students.

We infer the student social network from granular, session-level wifi data, and use the `igraph` package for R for social network analysis (Csardi & Nepusz, 2006). This network is represented by a graph of wifi-using students, connected by shared wifi usage. Edges are defined as concurrent connections via the same Access Point, and are weighted by the multiplicity of co-sessions, which is intended to proxy the intensity of social relationships. We calculate the weighted average of peer attributes. Students who spend a significant amount of time "together" in this manner are likely to be either friends or partners for course projects and study groups. Table 3.5 summarizes the student-level characteristics for Fall 2008. Results are based on Fall 2008, but the method can be extended to a dynamic graph of student-semesters.

This network is represented by graph $G = (V, E)$, where V represents the set of wifi using students. Edges, E, are defined by access to the same Access Point within the same five-minute time period, with the multiplicity of such concurrent periods serving as edge weight. Among 6,052 students, there are 3,073 wifi-using students and 1,178,400 edges connecting those students. E has a median weight of 11, with a minimum and maximum of 1 and 13,894 respectively.

Figure 3.3 shows the percentage of time a student spends, on average with his Nth friend. Note that students can spend time with multiple "friends" at once, so percentages do not need to add up to 100%. This figure suggests that the first 1-3 “friends” are most relevant in determining peer effect, with decreasing marginal time spent with all remaining “friends”. Percentage of time with Nth friend is comparable across all curricular years.

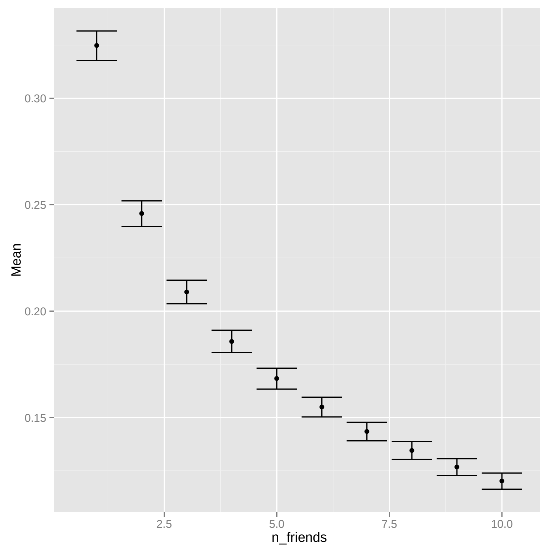


Figure 3.3: Percentage of Time over Neighbor Strength.

Table 3.4: Correlation Matrix—First Differences

	ΔNo. Courses	ΔGrade Points	ΔHours	ΔMegabytes	Application Score	Cohort
ΔNo. Courses	1					
ΔGrade Points	0.965	1				
ΔHours	0.128	0.138	1			
ΔMegabytes	0.0368	0.0330	0.470	1		
Application Score	-0.0252	-0.0286	-0.00366	-0.0173	1	
Cohort	-0.0229	-0.0355	-0.0157	-0.00329	0.102	1

Table 3.5: Social Network Descriptive Statistics (Fall 2008)

Statistic	Mean	Median	St. Dev.	Min	Max	N
Grade Points	57.1	57.0	35.0	10.0	646.0	3,327
No. Courses	52.6	53	26.7	10	332	2,376
Hours	65.7	60.0	28.7	37.0	646.0	3,327
Megabytes	61.6	58	19.7	37	332	2,376
Application Score	4.3	5	2.4	1	35	3,327
Cohort	4.1	4	1.9	1	22	2,376
No. Self Bins	90.4	38.7	126.4	0.01	1,255.5	3,327
Tot. Common Bins	8,108.7	1,643.1	21,533.7	0.02	370,134.6	3,327
Top 3 Neighbors Application Score	149.1	149.0	12.4	89.0	192.3	3,177
Top 3 Neighbors No. Courses	4.4	4.7	1.2	1.0	13.7	3,304
Top 3 Neighbors Grade Points	59.8	60.3	18.1	11.0	246.0	3,304
Top 3 Neighbors Grade Points*	66.5	65.3	12.9	41.3	255.0	3,304
Top 3 Neighbors No. Courses (Lagged)	4.4	4.3	1.2	1.0	15.0	3,175
Top 3 Neighbors Grade Points (Lagged)	57.2	57.3	17.5	10.0	213.0	3,175
Top 3 Neighbors Grade Points* (Lagged)	63.9	62.5	12.6	37.0	216.0	3,175

3.4 Discussion

It is important to recognize that the social network used in this work is a proxy for the true social network that exists among students. Thus, it is difficult to identify which relationships constitute true social influence, and which are spurious. This may be compared to the typical Type I and II errors in statistics. First, the social network may over-report student relationships where there is in fact no relationship (Type I False Positive error). Second, true interactions not involving wifi usage are not reported (Type II False Negative error).

Chapter 4

ICT Effects: Methodology & Model

4.1 Overview

The goal of the first part of this thesis is to estimate the effect of ICT, specifically on-campus wifi usage and laptop ownership as proxied by on-campus wifi usage, on student academic performance as measured by grades. This section establishes the methodologies, and models to estimate the ICT effect. This is important because ICT is an important input to the education production function.

4.2 Methodology

Our identification strategy is to use Ordinary Least Squares with First Differences (OLS-FD) to control for student-specific time-constant unobserved effects. OLS-FD controls for student-specific time-constant unobserved effects, student and semester fixed effects, but only for wifi users. We extend this methodology with a Dynamic Propensity Score Matching model that allows us extend the analysis to non-wifi users, estimate a laptop treatment effect, and corroborate OLS-FD results. Matching is ideal for how it reduces bias in observational (non-random) data, as we have, and also reduce model dependence.

4.2.1 Ordinary Least Squares–First Differences

We use Ordinary Least Squares with First Differences to estimate the effect of ICT (wifi) usage on student performance. In this technique, we take first differences between time periods, and regress on the differences, thus subtracting away student-specific time-constant unobserved effects.

4.2.2 Propensity Score Matching

OLS-FD is a great place to start, but only includes observations which have reported wifi usage. We extend the OLS-FD using Dynamic Propensity Score Matching, which has several advantages. First, it allows us to define an ICT treatment effect to account for students without wifi usage. Second, it allows us to argue more strongly for a causal result, since matching is able to simulate the experiment of random assignment to treatment and control groups. Finally, matching reduces model dependence, which is helpful when one cannot control for all desired covariates, as in our case.

Matching is a general, non-parametric procedure for reducing bias and model dependence in observational studies, and is therefore always preferable to unmatched analysis (Rubin, 1973), and the propensity score is used to reduce covariate dimensionality (Rosenbaum & Rubin, 1983). Dynamic PSM is a natural extension of PSM to panel data (Young, 2008), and is performed by matching separately for each semester and then combining semesters to form a matched panel. In this way, treated units always match to control units in the same semester, effectively controlling for semester. There are some caveats here as discussed in Young (2008, p. 91). Careful definition of the treatment variable can differentiate the Wifi Effect from the Laptop Effect. We assume that at least one wifi session constitutes ownership of a laptop.

Table 4.1: Propensity Score Matching Treatment Variables

	Wifi (T_W)	Laptop (T_L)	ICT (T_I)
Control ($T = 0$)	1^{st} Quintile Wifi Users	Non-Wifi Users	Non-Wifi Users
Treatment ($T = 1$)	5^{th} Quintile Wifi Users	1^{st} Quintile Wifi Users	5^{th} Quintile Wifi Users

Results are similar for first and tenth deciles.

PSM involves estimation of the propensity score specification that satisfies the balancing property—that is, the specification for which treatment and control groups have similar distributions after matching (Rosenbaum & Rubin, 1983). To facilitate comparison with OLS-FD results, we use all the covariates in Table 4.2 (except semester) in the propensity score specification.

Balance is evaluated in terms of percent bias reduction. A common, but incorrect, way to test the balancing property is with a difference-in-means t-test, to see if the treated and control groups are similar at given values of the propensity score. Imai, King, and Stuart (2008) address this issue in detail. (Imai et al., 2008, p. 495). Balancing performance is presented in Section 5.2.

The Laptop Effect treatment group is defined to be the lower quintile of usage, and the control group is all users with zero wifi usage, those that either do not own a laptop, do not bring it to campus, or do not connect it to wifi. The Wifi Effect treatment group is defined by the upper quintile of wifi usage—this group clearly engages in bona fide wifi use. The control group is the group of laptop owners as defined before. In this way we differentiate bona fide wifi users from those who choose not to use wifi (proxied by those who use a trivial amount), and those who cannot use wifi (because they most likely do not own a laptop). The Total Effect is obtained by comparing the wifi treatment group to the laptop control group, to capture the full effect of obtaining a laptop *and* using it for wifi. Table 4.1 outlines three binary treatment variables to measure these effects.

4.3 Model

As defined in Table 4.2, let $P_{i,t}$ represent student i's performance in semester t, and let $W_{i,t}$ represent student i's wifi usage in semester t. Let $\boldsymbol{Z}_{i,t}$ be a vector of student, semester, and student-semester covariates. The standard OLS model is given in Equation 4.1.

$$P_{i,t} = \beta_0 + W_{i,t}\beta_W + \boldsymbol{Z}'_{i,t}\boldsymbol{\beta_Z} + U_i\beta_U + \epsilon_{i,t} \tag{4.1}$$

Table 4.2: Model Variables

	Model Variable	Description
$P_{i,t}$	$GP_{i,t}$	Grade Points
	$NC_{i,t}$	No. Courses
$W_{i,t}$	$HR_{i,t}$	Hours Online
	$MB_{i,t}$	Megabytes Transferred
	$BW_{i,t}$	Bandwidth (MB/HR)
$\boldsymbol{Z}_{i,t}$	AS_i	Application Score
	$\boldsymbol{CH}_i$	Cohort Dummies
	$\boldsymbol{CY}_{i,t}$	Curricular Year Dummies
	$\boldsymbol{MJ}_i$	Major Dummies
	$\boldsymbol{SM}_t$	Semester Dummies
	U_i	Unobserved Effects

We use Application Score, with which the student applied to FEUP, to control for prior student performance, and indirectly for students' attributes, e.g. aptitude, socioeconomic status, gender, family composition, and so forth. We sought socioeconomic survey results, as collected by higher education institutions in Portugal, but could not match them with a representative subset of the data.

We use Cohort, or the year in which a student started as a first-year student, to control for differences in the overall performance of cohorts. Dummy variables are included for Curricular Year, Major, and Semester to control, respectively, for variations in grades and wifi usage between various curricular years, differences in major difficulty or grading policies, and differences through absolute time.

4.3.1 Wifi Effect

First Differences Taking First Differences (FD) eliminates student-specific time-constant effects, U_i, and yields:

$$\Delta P_{i,t} = \Delta W_{i,t}\beta_W + \boldsymbol{Z'_{i,t}\beta_Z} + \Delta\epsilon_{i,t} \tag{4.2}$$

where $\Delta P_{i,t} = P_{i,t} - P_{i,t-1}$, and so forth. The wifi effect is given by β_W, which shows how changes in wifi usage relate to changes in performance. If wifi usage is positively correlated with academic performance, then $\beta_W > 0$. So, the null hypothesis is $H_0 : \beta_w = 0$.

Day and Night Usage We want to consider the effect of wifi usage on performance, separated by day and night. Let $D_{i,t}$ be the total wifi usage of student i in semester t that occurred between the hours of 8 a.m. to 8 p.m. (daytime). Let $N_{i,t}$ be the total wifi usage for student i during semester t that occurred between 8 p.m. to 8 a.m. (nighttime). Thus, $D_{i,t} + N_{i,t} = W_{i,t}$. This model is a simple adaptation of Equation 4.2.

$$\Delta P_{i,t} = \beta_0 + \Delta D_{i,t}\beta_d + \Delta N_{i,t}\beta_n + \boldsymbol{Z'_{i,t}\beta_z} + \Delta\epsilon_{i,t} \tag{4.3}$$

We hypothesize that wifi usage, both during the day and at night, is productive, so we posit $H_0 : \beta_d > 0,\ \beta_n > 0$. Further, we suppose that daytime usage is likely to be more productive than nighttime usage since most work is done during the day, so we hypothesize $H_0 : \beta_d > \beta_n$.

Curricular Year In this model we interact curricular year with wifi usage to control for differences in the effect of wifi for each curricular year; for instance, we expect graduating master's students to be more productive than first-year undergraduates in the way they use wifi. We interact curricular year dummies $\boldsymbol{Y_{i,t}}$, with usage to control for different wifi usage between different curricular years:

$$\Delta P_{i,t} = \beta_0 + \Delta W_{i,t}\beta_w + \Delta W_{i,t}\boldsymbol{Y'_{i,t}\beta_y} + \boldsymbol{Z'_{i,t}\beta_z} + \Delta\epsilon_{i,t} \tag{4.4}$$

We believe $H_0 : \boldsymbol{\beta_y} \neq \mathbf{0}$, that the effect of wifi differs according to curricular year. Further, we posit $H_0 : \beta_{y_5} > \beta_{y_4} > \ldots > \beta_{y_1}$, which would indicate that student wifi productivity is commensurate with maturity.

Major In this model we control for differences among majors. We interact major with usage to account for differences in the way different majors use wifi. Let $\boldsymbol{M_i}$ denote the five major dummy variables.

$$\Delta P_{i,t} = \beta_0 + \Delta W_{i,t}\beta_w + \Delta W_{i,t}\boldsymbol{M_i'}\boldsymbol{\beta_m} + \boldsymbol{Z_{i,t}'}\boldsymbol{\beta_z} + \Delta\epsilon_{i,t} \tag{4.5}$$

The research hypothesis under this model is $H_0 : \boldsymbol{\beta_m} \neq \mathbf{0}$. We believe that some majors with an affinity for ICT, such as Computer Science, may use wifi more productively than other majors.

Bandwidth In this model we control for bandwidth, B, or the usage ratio Megabytes/Hours.

$$\Delta P_{i,t} = \beta_0 + \Delta B_{i,t}\beta_W + \boldsymbol{Z_{i,t}'}\boldsymbol{\beta_z} + \Delta\epsilon_{i,t} \tag{4.6}$$

Time on Campus Finally, in this model we control for time on campus, C, which is calculated as the average of the time in hours between the start of the first session and the end of the last session each day.

$$\Delta P_{i,t} = \beta_0 + \Delta C_{i,t}\beta_W + \boldsymbol{Z_{i,t}'}\boldsymbol{\beta_z} + \Delta\epsilon_{i,t} \tag{4.7}$$

4.3.2 Laptop Effect

Propensity Score Matching Model

Equation 4.8 gives the general propensity score specification, where $T_{i,t}$ is one of the three treatment variables defined in Table 4.1. Equation 4.8 is estimated using a probit.

$$T_{i,t} = \beta_0 + \boldsymbol{Z}_{i,t}'\boldsymbol{\beta_Z} + \nu_{i,t} \tag{4.8}$$

4.4 Empirical Strategy

We set up two models to estimate the Wifi Effect on student performance, to characterize that effect by usage patterns and student covariates, and to differentiate wifi, and laptop effects.

First, we perform Ordinary Least Squares (OLS) regressions to estimate the baseline Wifi Effect on Performance, with First Differences (FD) to control for unobserved time-constant heterogeneity among students. For robustness and since some covariates are missing for a subset of the students, I present results with and without these extra covariates.

Second, we run dynamic Propensity Score Matching (PSM) models to estimate the average treatment effect on the treated (ATT), which replicates the OLS result (though coefficients are not directly comparable).

As noted in Chapter 3, the data have some coverage gaps in covariates. For instance, Application Score is missing for 55% of students. To increase the robustness of these results, we run regressions with and without covariates, to see if the model with additional observations displays the same qualitative results as the model with fewer observations. Therefore, a total of eight regressions are provided for each of the models given above.

Variables are standardized independently for each model, so coefficients should be interpreted as the percent of a standard deviation change in $P_{i,t}$ given a standard deviation change in the independent variable.

Chapter 5

ICT Effects: Results

5.1 Results with Ordinary Least Squares

5.1.1 Wifi Effect

We perform Ordinary Least Squares (OLS) regressions using First Differences (FD) to control for observed time-constant heterogeneity among students. Since some covariates are missing for a subset of the students, we present results with and without these extra covariates as a robustness check. Interesting results for Q2 are found for related models. Notably, we find evidence that more mature students (e.g., fifth-year Masters students) are among the most productive wifi users. We also find daytime usage to be more productive than nighttime usage, and weak heterogeneity of wifi effects across majors.

Coefficients are normalized by the standard deviation over observations included in the regression, so coefficients may be interpreted as percent of a standard deviation change in $P_{i,t}$ given a standard deviation change in the independent variable. For example, a coefficient of $0.123 = 12.3\%$ is understood to mean that an increase of one standard deviation in the independent variable (x) is associated with a 12.3% of a standard deviation change in the dependent variable (y).

Table 5.1: Wifi Effect on Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	ΔNo. Courses	ΔNo. Courses	ΔNo. Courses	ΔNo. Courses	ΔGrade Points	ΔGrade Points	ΔGrade Points	ΔGrade Points	ΔGrade Points*	ΔGrade Points*	ΔGrade Points*	ΔGrade Points*
ΔHours	0.128***	0.150***			0.138***	0.162***			0.113***	0.144***		
	(0.0150)	(0.0161)			(0.0156)	(0.0168)			(0.0144)	(0.0174)		
ΔMegabytes			0.0368***	0.0336**			0.0330**	0.0266*			0.0311**	0.0292*
			(0.0137)	(0.0155)			(0.0141)	(0.0157)			(0.0131)	(0.0168)
Application Score		-0.0303***		-0.0320***		-0.0452***		-0.0472***		-0.0562***		-0.0579***
		(0.0112)		(0.0111)		(0.0115)		(0.0114)		(0.0123)		(0.0122)
Constant	-0.164***	0.165	-0.138***	0.123	-0.136***	0.121	-0.108***	0.0772	-0.120***	0.0596	-0.0978***	0.0201
	(0.0119)	(0.143)	(0.0112)	(0.141)	(0.0120)	(0.140)	(0.0112)	(0.139)	(0.0121)	(0.131)	(0.0115)	(0.128)
Cohort Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Curricular Year Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Major Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Semester Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N (student-semesters)	6425	5337	6425	5337	6425	5337	6425	5337	6425	5337	6425	5337
N (students)	3030	2421	3030	2421	3030	2421	3030	2421	3030	2421	3030	2421

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Main Result

OLS-FD results show a positive, statistically significant correlation, although this effect is not economically large[1]. Table 5.1 shows regressions for ΔGrade Points and ΔNo. Courses on both ΔHours and ΔMegabytes, both with and without covariates (even and odd columns, respectively). The results show a Wifi Effect of 12.8–16.2% as measured by Hours, and 2.7–3.7% as measured by Megabytes. All coefficients are statistically significant at the 10% level, and most are significant at the 1% level. Coefficients for ΔGrade Points are larger than ΔNo. Courses when measuring by ΔHours, and smaller measuring by ΔMegabytes.

Both ΔHours and ΔMegabytes have a statistically significant, positive effect on $P_{i,t}$. There is an 13.8% effect of ΔHours on ΔGrade Points, and an even greater effect when controlling for student-specific time-varying factors (16.2%). ΔNo. Courses shows similar effects of 12.8% and 15.0%, respectively. There is a 3.3% and 3.7% effect of ΔMegabytes on ΔGrade Points and ΔNo. Courses, respectively, reaffirming the positive relationship seen with ΔHours, although significance is slightly reduced. When controlling for covariates, there is a smaller, but statistically significant effect of 2.7% and 3.4% for ΔGrade Points and ΔNo. Courses, respectively.

Application Score is significant and negative, with similar signs and magnitudes for both ΔGrade Points and ΔNo. Courses (being somewhat smaller for the latter). This is reasonable, because while prior performance is positively correlated with current performance, it is negatively correlated with a student's marginal capacity to change performance.

It is not clear why the magnitude of effect for ΔHours is larger than for ΔMegabytes. First, it is possible that ΔHours is partially a proxy for time spent on campus. This can be controlled for by defining a time-on-campus variable as the average elapsed time between the first login and the last logout for reasonable daily arrival and departure times. Second, it may be that time-intensive users are more productive than data-intensive users. The

[1] This is consistent with Sosin *et al.* (2004) cited in Ben Youssef & Dahmani, 2008, p. 48.

Megabytes/Hours ratio, which represents throughput, can be used in this case to differentiate "hot" vs. "cold" wifi usage (McLuhan, 1964).[2] One expects this ratio to be positive for data-intensive users, and negative for time-intensive users.

Day and Night Wifi Usage

We examine the effect of $W_{i,t}$ on $P_{i,t}$ when W is split into day (D) and night (N) usage. Table 5.2 shows regressions of ΔHours (Day) and ΔHours (Night) on $P_{i,t}$. Note first that ΔHours (Day) is almost identical to ΔHours from Table 5.1. ΔHours (Night) is also positive, but much lower in magnitude and weaker in significance. All other coefficients are statistically significant at the 1% level. We see that ΔHours (Day) has a substantial 17.0% and 15.7% effect on ΔGrade Points and ΔNo. Courses respectively, and controlling for $\boldsymbol{Z_{i,t}}$ decreases this effect, with 15.3% and 14.1% effects respectively. We find it notable that coefficients for $\boldsymbol{Z_{i,t}}$ remain substantially the same between models, and thereby gain confidence in this analysis.

Day/Night results provide a very unique insight, that the positive effect for ΔHours that we saw earlier actually results from daytime usage, when class is in session. It seems reasonable that students who work during the day appear to be more productive than their night owl colleagues. Nighttime usage is also positive, but of smaller significance, or insignificant.

Table 5.2 also shows day/night megabyte usage. The results for ΔMegabytes (Day) follow ΔMegabytes from Table 5.1, and we observe again that daytime usage carries the overall positive effect seen in Tables 5.1. ΔNighttime Megabytes is statistically zero when controlling for $\boldsymbol{Z_{i,t}}$. Signs and magnitudes of control variables in $\boldsymbol{Z_{i,t}}$ follow first difference results from Section 5.1.1. These results are further support first difference results and day/night usage by hours.

[2] The data, unfortunately, provide no other information on how students use wifi.

Table 5.2: Wifi Effect on Performance by Day/Night

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	ΔNo. Courses	ΔNo. Courses	ΔNo. Courses	ΔNo. Courses	ΔGrade Points	ΔGrade Points	ΔGrade Points	ΔGrade Points	ΔGrade Points*	ΔGrade Points*	ΔGrade Points*	ΔGrade Points*
ΔHours (Day)	0.106***	0.137***			0.120***	0.151***			0.106***	0.133***		
	(0.0161)	(0.0169)			(0.0169)	(0.0179)			(0.0160)	(0.0190)		
ΔHours (Night)	0.0397**	0.0265			0.0334**	0.0232			0.0167	0.0236		
	(0.0163)	(0.0173)			(0.0166)	(0.0174)			(0.0159)	(0.0186)		
ΔMegabytes (Day)			0.0270*	0.0315*			0.0233	0.0234			0.0237	0.0214
			(0.0148)	(0.0171)			(0.0154)	(0.0175)			(0.0145)	(0.0188)
ΔMegabytes (Night)			0.0172	0.00486			0.0168	0.00620			0.0132	0.0135
			(0.0133)	(0.0154)			(0.0130)	(0.0151)			(0.0127)	(0.0172)
Application Score		-0.0306***		-0.0320***		-0.0456***		-0.0472***		-0.0566***		-0.0579***
		(0.0112)		(0.0111)		(0.0115)		(0.0114)		(0.0123)		(0.0122)
Constant	-0.164***	0.164	-0.138***	0.122	-0.136***	0.120	-0.107***	0.0770	-0.122***	0.0590	-0.0977***	0.0215
	(0.0120)	(0.143)	(0.0112)	(0.141)	(0.0121)	(0.140)	(0.0112)	(0.139)	(0.0123)	(0.131)	(0.0115)	(0.129)
Cohort Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Curricular Year Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Major Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Semester Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N (student-semesters)	6425	5337	6425	5337	6425	5337	6425	5337	6425	5337	6425	5337
N (students)	3030	2421	3030	2421	3030	2421	3030	2421	3030	2421	3030	2421

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Curricular Year

Using the original first difference model, we regress on the interaction of $W_{i,t}$ and curricular year, giving a different slope for each curricular year with the first curricular (freshman) year as the base case[3]. This shows us how the impact of $W_{i,t}$ changes as a student progresses through their degree. Table 5.3 shows these regressions.

We see little statistical significance in the first year; however, coefficients steadily increase in magnitude and significance to show, in curricular year 5, a 15.7% and 14.2% effect on ΔGrade Points and ΔNo. Courses without $\boldsymbol{Z_{i,t}}$, and respective 16.2% and 14.4% effects with $\boldsymbol{Z_{i,t}}$. Coefficients on $\boldsymbol{Z_{i,t}}$ continue to follow the magnitude and sign of other regression sets. As before, ΔGrade Points is generally larger than ΔNo. Courses.

This is a very interesting result. We see that ΔHours is zero (or small) for all students in their first curricular year. We observe a trend that coefficients increase from zero to a very robust 14.2%–16.2% effect in curricular year five. This suggests that the wifi is increasingly useful as students gain more maturity in their academic program. It is notable that, when controlling for $\boldsymbol{Z_{i,t}}$, students in curricular years 4 and 5 have such a significantly positive relationship between wifi usage and grades. Since students finish their licentiate after 3 years, we see that the internet is most useful for master's students. This is consistent with the hypothesis that the utility of the internet in improving grades is correlated with student maturity.

Next, we consider megabyte usage interacted against curricular year, shown in Table 5.3. In contrast with hours regressions, ΔMegabytes actually shows a negative effect on grades at the 1% level when controlling for $\boldsymbol{Z_{i,t}}$ in curricular year 1. Again, there is an increasing trend from a negative first-year effect, to a positive and significant effect for master's students in curricular years 4 and 5. As in other regressions on megabyte usage, coefficients are generally

[3]Actually, the base case is a trivial number of observations without curricular year defined, for which curricular year 1 is not significantly different except in column (2).

Table 5.3: Wifi Effect on Performance by Curricular Year

	(1) ΔNo. Courses	(2) ΔNo. Courses	(3) ΔNo. Courses	(4) ΔNo. Courses	(5) ΔGrade Points	(6) ΔGrade Points	(7) ΔGrade Points	(8) ΔGrade Points	(9) ΔGrade Points*	(10) ΔGrade Points*	(11) ΔGrade Points*	(12) ΔGrade Points*
ΔHours * CY_1	-0.00365	0.0148			-0.0154	0.0129			-0.0226**	0.00735		
	(0.0121)	(0.0142)			(0.0122)	(0.0144)			(0.00900)	(0.0123)		
ΔHours * CY_2	0.0584***	0.0383***			0.0563***	0.0367***			0.0374***	0.0249***		
	(0.0128)	(0.0135)			(0.0124)	(0.0128)			(0.00855)	(0.00955)		
ΔHours * CY_3	0.0532***	0.0352***			0.0532***	0.0368***			0.0404***	0.0229**		
	(0.00961)	(0.0113)			(0.00893)	(0.0105)			(0.00794)	(0.00938)		
ΔHours * CY_4	0.0820***	0.0936***			0.0898***	0.0954***			0.0805***	0.0991***		
	(0.0122)	(0.0154)			(0.0127)	(0.0157)			(0.0131)	(0.0178)		
ΔHours * CY_5	0.0748***	0.115***			0.0894***	0.132***			0.0769***	0.117***		
	(0.0175)	(0.0179)			(0.0189)	(0.0198)			(0.0175)	(0.0209)		
ΔMegabytes * CY_1			-0.0109	0.00327			-0.0180*	0.00363			-0.0190*	0.00440
			(0.0101)	(0.0108)			(0.0108)	(0.0110)			(0.0101)	(0.0108)
ΔMegabytes * CY_2			0.0250	-0.000581			0.0218	-0.00612			0.0146*	-0.0104
			(0.0155)	(0.0135)			(0.0142)	(0.0117)			(0.00819)	(0.00651)
ΔMegabytes * CY_3			0.0148	0.00137			0.01000	-0.00677			0.00826	-0.0117
			(0.0105)	(0.0135)			(0.00946)	(0.0126)			(0.00780)	(0.0121)
ΔMegabytes * CY_4			0.0338***	0.0417***			0.0344***	0.0384***			0.0361***	0.0471***
			(0.0119)	(0.0143)			(0.0123)	(0.0143)			(0.0137)	(0.0180)
ΔMegabytes * CY_5			0.00714	0.0148			0.00769	0.0149			0.00889	0.0167
			(0.0151)	(0.0159)			(0.0165)	(0.0173)			(0.0146)	(0.0180)
Application Score		-0.0312***		-0.0321***		-0.0464***		-0.0475***		-0.0573***		-0.0582***
		(0.0113)		(0.0111)		(0.0117)		(0.0115)		(0.0124)		(0.0122)
Constant	-0.163***	0.182	-0.139***	0.125	-0.134***	0.142	-0.107***	0.0786	-0.118***	0.0846	-0.0972***	0.0215
	(0.0119)	(0.143)	(0.0113)	(0.142)	(0.0120)	(0.141)	(0.0113)	(0.139)	(0.0122)	(0.132)	(0.0116)	(0.129)
Cohort Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Curricular Year Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Major Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Semester Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N (student-semesters)	6425	5337	6425	5337	6425	5337	6425	5337	6425	5337	6425	5337
N (students)	3030	2421	3030	2421	3030	2421	3030	2421	3030	2421	3030	2421

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

slightly higher for ΔGrade Points than for ΔNo. Courses. Signs and magnitudes for $\boldsymbol{Z_{i,t}}$ continue to follow previous results.

We see some very interesting patterns. For both hours and megabytes, the positive effect is most strongly evident at the master's level, although the increasing effect is visible for licentiate students also. In all cases, the choice of $\boldsymbol{Z_{i,t}}$ adequately controls for student specific time varying effects and we generally see an increase in statistical significance when adding $\boldsymbol{Z_{i,t}}$ to a regression.

Major

We interact wifi usage with student major in the last model. Table 5.4 shows the results for interactions with Hours. These results are quite unique compared to the previous three models. In the base case (Other Majors) there is a substantial, statistically significant effect of 21.4% and 25.0% effect of Grade Points without and with covariates, and 19.6% and 21.5% for No. Courses. The Computer Science interaction is zero, so these high effects hold for this major (as also for Electrical and Mechanical Engineering measuring by No. Courses). Major interaction dummies on the remaining majors yield an effect from 15.0%–17.6% by Grade Points and 14.9%–15.5% for No. Courses. No interaction is significant when including covariates. The interaction variables show an ordering of the wifi productivity among majors. Computer Science seems to use wifi best, which would be consistent with expectations, although the remaining majors do not have a clear ordering for both Grade Points and No. Courses.

Table 5.4 also shows results for interactions with Megabytes. Without covariates, only Computer Science shows a positive effect, and all other majors are statistically zero. Things get more interesting for the results with covariates. The main interaction is negative and significant for both Grade Points (at -8.4%) and No. Courses (at -7.8%). All other majors, besides Chemical Engineering, show an improvement, from -4.8% to 0.0% by Grade Points

Table 5.4: Wifi Effect on Performance by Major

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	ΔNo. Courses	ΔNo. Courses	ΔNo. Courses	ΔNo. Courses	ΔGrade Points	ΔGrade Points	ΔGrade Points	ΔGrade Points	ΔGrade Points*	ΔGrade Points*	ΔGrade Points*	ΔGrade Points*
ΔHours	0.196***	0.215***			0.214***	0.250***			0.192***	0.245***		
	(0.0522)	(0.0741)			(0.0556)	(0.0759)			(0.0483)	(0.0709)		
ΔHours * ECE	-0.0560	-0.0602			-0.0638*	-0.0801			-0.0597*	-0.0833*		
	(0.0354)	(0.0495)			(0.0377)	(0.0509)			(0.0332)	(0.0485)		
ΔHours * CS	0.0118	-0.0117			0.00976	-0.0151			-0.00109	-0.0157		
	(0.0275)	(0.0382)			(0.0293)	(0.0393)			(0.0258)	(0.0376)		
ΔHours * ME	-0.0354	-0.0168			-0.0472*	-0.0411			-0.0505**	-0.0556		
	(0.0278)	(0.0370)			(0.0283)	(0.0370)			(0.0251)	(0.0357)		
ΔHours * CHE	-0.0466**	-0.0338			-0.0397*	-0.0358			-0.0467***	-0.0607**		
	(0.0195)	(0.0274)			(0.0207)	(0.0281)			(0.0172)	(0.0260)		
ΔHours * CE	-0.0409**	-0.0222			-0.0377*	-0.0207			-0.0310	-0.0132		
	(0.0205)	(0.0276)			(0.0220)	(0.0295)			(0.0206)	(0.0279)		
ΔMegabytes			-0.0334	-0.0778**			-0.0339	-0.0838*			-0.0398	-0.0914*
			(0.0410)	(0.0363)			(0.0523)	(0.0459)			(0.0547)	(0.0550)
ΔMegabytes * ECE			0.0313	0.0574**			0.0296	0.0552*			0.0390	0.0673*
			(0.0291)	(0.0274)			(0.0353)	(0.0321)			(0.0363)	(0.0375)
ΔMegabytes * CS			0.0641**	0.0783***			0.0633**	0.0816***			0.0567*	0.0815**
			(0.0259)	(0.0239)			(0.0316)	(0.0290)			(0.0327)	(0.0345)
ΔMegabytes * ME			0.0264	0.0534**			0.0202	0.0449**			0.0207	0.0421*
			(0.0210)	(0.0209)			(0.0235)	(0.0221)			(0.0229)	(0.0236)
ΔMegabytes * CHE			0.00894	0.0232			0.0129	0.0284			0.0187	0.0365*
			(0.0165)	(0.0181)			(0.0188)	(0.0200)			(0.0165)	(0.0192)
ΔMegabytes * CE			0.0168	0.0344***			0.0148	0.0359***			0.0160	0.0444***
			(0.0122)	(0.0106)			(0.0146)	(0.0129)			(0.0165)	(0.0155)
Application Score		-0.0303***		-0.0317***		-0.0456***		-0.0472***		-0.0562***		-0.0578***
		(0.0112)		(0.0111)		(0.0115)		(0.0114)		(0.0123)		(0.0122)
Constant	-0.162***	0.152	-0.138***	0.127	-0.134***	0.106	-0.107***	0.0818	-0.119***	0.0379	-0.0972***	0.0248
	(0.0118)	(0.142)	(0.0112)	(0.141)	(0.0119)	(0.140)	(0.0112)	(0.139)	(0.0120)	(0.130)	(0.0115)	(0.128)
Cohort Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Curricular Year Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Major Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Semester Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N (student-semesters)	6425	5337	6425	5337	6425	5337	6425	5337	6425	5337	6425	5337
N (students)	3030	2421	3030	2421	3030	2421	3030	2421	3030	2421	3030	2421

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

and -4.3% to 0.0% by No. Courses.

Megabytes results are particularly interesting for their contrast to Hours results. While the academic productivity of wifi usage seems to change across majors by Hours, such differences drop off significantly, and even become negative, when wifi is measured by Megabytes. In the next section, we examine the relationship between Hours and Megabytes by regressing on the bandwidth, or usage ratio Hours/Megabytes.

5.2 Results with Dynamic Propensity Score Matching

5.2.1 Wifi & Laptop Effect

As discussed in 4.2.2, we want to see if the propensity score specification adequately balances treatment and control group observations—that is, if when matching by propensity score, treatment and control observations are more similar to after matching than before matching. Figure 5.1 shows percent bias reduction by covariate for the Grade Points–Hours model, relative to largest absolute bias, and is negative where matched bias > unmatched bias. One can see that percent bias reduction is quite good in most cases; for example, bias in Application Score reduces for all but one semester for all treatment effects. Where balancing technically worsens (percent bias reduction is negative), it is never large in an absolute sense. In most cases we see very good balancing, and with an slight increase only where bias was already very low[4].

Table 5.5 shows the total number of observations and the number of observations selected by the matching algorithm. Only matched observations are used in the analysis.

The PSM results show a statistically significant relationship for Wifi and Total Effects, and a partially significant result for Laptop Effect. These results are presented in Figure

[4]Young (2008) shows similar balancing performance, with a reduction in bias for most covariates with an occasional increase)

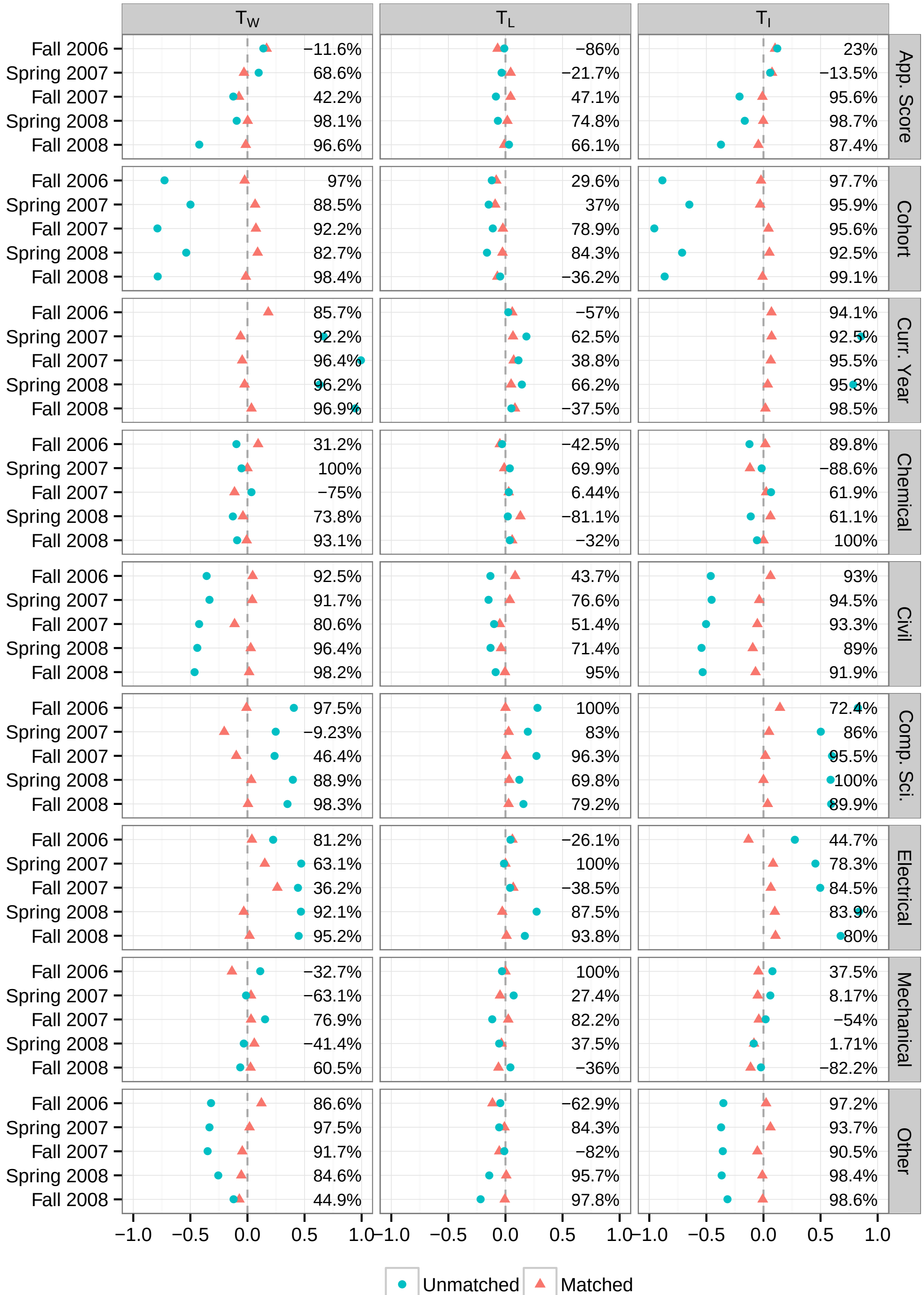


Figure 5.1: Percent Bias Reduction (Relative to Highest Value)

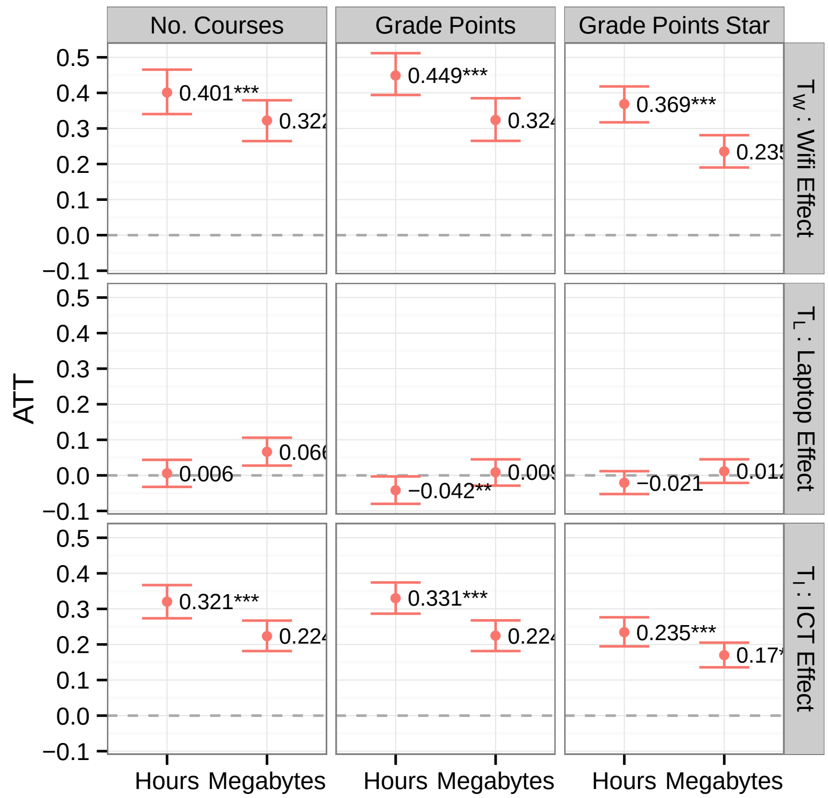


Figure 5.2: Propensity Score Matching Results: Average Treatment Effect on the Treated (ATT) by Treatment Effect, Performance, and Wifi

Table 5.5: No. Observations

Statistic	Median
Control_All	1,898
Control_Matched	811
Control_Unmatched	877
Control_Discarded	210
Treated_All	2,059
Treated_Matched	2,032
Treated_Unmatched	0
Treated_Discarded	27

5.2. The Wifi Effect is 37.9–41.4% by Hours and 23.5–25.2% by Megabytes, all statistically significant at the 1% level. The Laptop Effect is statistically zero by Hours, and slightly positive by Megabytes, whereas the Total Effect is large, positive, and statistically significant. Comparing Total Effect to Wifi Effect, Hours is relatively smaller, and Megabytes is relatively larger. Note that these coefficients should not be directly compared with OLS-FD results since they are measuring the effect of dichotomous treatment, but Fixed Effects regressions on the identical variables and observations yield similar magnitude and significance.

We estimate the average treatment effect on the treated (ATT) using Ho, Imai, King, and Stuart (2011) and Imai, King, and Lau (2007). Figure 5.2 shows the ATT by performance, treatment, and wifi. T_W dominates the positive effect, while T_L is statistically zero (by Hours; T_L is slightly positive by Megabytes). Using a laptop in the absence of wifi does not significantly improve performance. Performance improves when the laptop is used to connect to wifi. Using the wifi network is what seems to have the positive association with grades; therefore, the total effect is somewhat similar to the wifi effect.

By hours, T_L is statistically zero, however by Megabytes T_L is 4.8% at 5% confidence for Grade Points, and 6.9% at 10% confidence for No. Courses. It is interesting to see that Megabytes has a positive and statistically significant effect, and as Megabyte usage is veritably also wifi usage, this likely reflects the bleeding of this choice of proxy into the Wifi Effect (whereas one may consistently believe that Hours does not indicate real wifi usage, for small levels of usage).

The Wifi Effect is much larger, and statistically significant at the 1% level. By Hours, T_W is 41.4% on Grade Points, and 37.9% on No. Courses. By Megabytes, T_W is 25.2% by Grade Points, and 23.5% by No. Courses. The effect is substantially larger by Hours than by Megabytes, as seen in OLS-FD results (Table 5.1). It is also slightly larger by Grade Points than by No. Courses.

5.3 Results & Discussion

Using both OLS and PSM, we conclude that wifi usage has a positive relationship with academic performance among FEUP students. This effect is generally larger for Hours online than Megabytes transferred. These results clarify the mechanics of the Wifi Effect, and partially address endogeneity. Using PSM, I show that laptop ownership (distinct from wifi usage) has a statistically zero (or nearly zero) relationship with grades when measured by Hours, and actual wifi usage (not mere laptop ownership) drives the observed positive effect. Table 5.6 summarize these reslts.

Table 5.6: Main Results (N=6425)

	ΔNumber of Courses	ΔGrade Points
ΔTotal Hours	0.150*** (0.0161)	0.162*** (0.0168)
ΔTotal Megabytes	0.0336** (0.0155)	0.0266* (0.0157)
ΔHours × CY 1	0.0148 (0.0142)	0.0129 (0.0144)
ΔHours × CY 2	0.0383*** (0.0135)	0.0367*** (0.0128)
ΔHours × CY 3	0.0352*** (0.0113)	0.0368*** (0.0105)
ΔHours × CY 4	0.0936*** (0.0154)	0.0954*** (0.0157)
ΔHours × CY 5	0.115*** (0.0179)	0.132*** (0.0198)
ΔMegabytes × CY 1	0.00327 (0.0108)	0.00363 (0.0110)
ΔMegabytes × CY 2	-0.000581 (0.0135)	-0.00612 (0.0117)
ΔMegabytes × CY 3	0.00137 (0.0135)	-0.00677 (0.0126)
ΔMegabytes × CY 4	0.0417*** (0.0143)	0.0384*** (0.0143)
ΔMegabytes × CY 5	0.0148 (0.0159)	0.0149 (0.0173)
ΔDay Hours	0.137*** (0.0169)	0.151*** (0.0179)
ΔNight Hours	0.0265 (0.0173)	0.0232 (0.0174)
ΔDay Megabytes	0.0315* (0.0171)	0.0234 (0.0175)
ΔNight Megabytes	0.00486 (0.0154)	0.00620 (0.0151)
Wifi Effect (Hours)	0.425***	0.389***
Wifi Effect (Megabytes)	0.270***	0.253***
Laptop Effect (Hours)	0.022	-0.003
Laptop Effect (Megabytes)	0.030	0.028
ICT Effect (Hours)	0.355***	0.318***
ICT Effect (Megabytes)	0.254***	0.245***

Some clear policy implications emanate from this work. First, it validates the widespread deployment of wifi in higher education. Second, it demonstrates that a first-year ICT orientation is in order since first-year students consistently misuse wifi. And third, it suggests that students may benefit from the ability to borrow laptops from the university library, since the benefits of laptop use accrue in wifi use and not so much in laptop ownership (non-wifi use).

Chapter 6

Peer Effects: Methodology & Model

6.1 Overview

As discussed in Chapter 1, peer effects are important in many disciplines. For example, marketers can harness an understanding of peer effects to design more effective advertising campaigns. Nonlinear peer effects are especially important in the context of education, and are particularly relevant for policy. The key question here is whether higher-performing students help or are hurt by their lower-performing peers, and vice versa. The former case would suggest a policy of grouping high- and low-performing students, whereas the latter case would suggest a policy of tracking or separating higher- and lower-performing students.

Social networks contain three principle sources of social correlation. First, it is well understood that "birds of a feather flock together"—a phenomenon known as homophily. Second, social correlation many result from peer influence in which a node in the network induces another node to adopt similar characteristics. Third, spurious, situational correlation is possible, such as correlation result from attending the same course but not interacting or sharing any latent characteristics. The goal of this work is to measure peer effects among students, and to differentiate heterogeneous peer effects among different types of students.

Homophily and peer influence can be viewed as opposite sides of the same coin: with homophily, nodes share a latent attribute which leads them to make a connection; with peer influence, nodes share a connection which leads them to adopt the same attributes (La Fond & Neville, 2010).

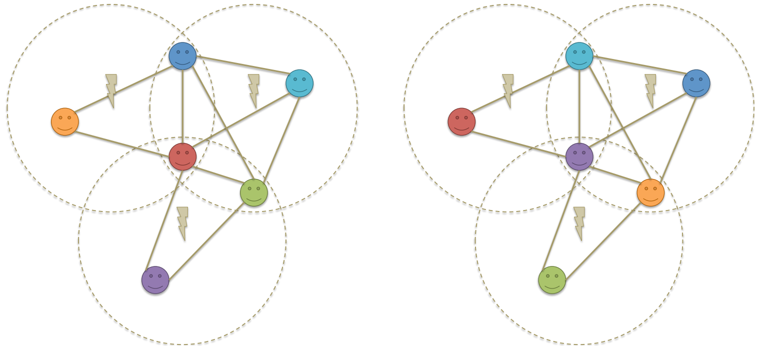
Our identification strategy is a randomization technique known as the Shuffle Test, which seeks to disentangle true peer influence from social correlation. In addition, every effort is made to control for known sources of situational correlation. Other methods will be discussed and employed for robustness.

6.2 Methodology

6.2.1 Shuffle Test

We use the shuffle test to disentangle the peer effect from other sources of social correlation. Randomization has been shown to be effective at identifying peer influence in the presence of homophily and confounding factors (Anagnostopoulos et al., 2008; Aral, Muchnik, & Sundararajan, 2009; Belo & Ferreira, 2012, 2013). The basic idea is to shuffle the social network in a manner that is orthogonal to the attributes under investigation. For example, if student performance and peer influence are independent from Access Point location, we can shuffle sessions among Access Points to break social ties without introducing misleading bias into the analysis. Since social links are defined by contemporaneous usage (same time and place), randomization by session Access Point location effectively breaks the social links among peer students, while retaining behavior explained by temporal usage patterns.

Figure 6.1 demonstrates the randomization methodology in a stylized diagram. In Figure 6.1a, we have the real social network for one instant in time, where individuals using the same access point at that instant are linked with an edge in the social network. Figure 6.1b shows one iteration of randomization, in which each the Access Point for this session is



(a) Original (b) Shuffled

Figure 6.1: Randomization Demonstration

shuffled across all users with a currently active session.

Assuming that the randomization is done "correctly" (to be discussed), we effectively argue that the randomization produces an alternate world, called a pseudosample, in which students have similar usage patterns and behavior (due to the homophily present in when and how students choose to connect to the wifi network) but zero peer influence, by construction (La Fond & Neville, 2010, p. 4). Any social correlation in the pseudosample cannot be due to peer influence; that is, the amount of peer influence can be estimated as the difference between the real social correlation and the randomized social correlation. Simulation over many pseudosamples yields a distribution for social correlation, and the difference from the true social correlation is the estimate of peer influence. A causal peer effect may thus be estimated as the difference between the actual peer effect and the mean randomized peer effect.

Notationally, true social correlation, or correlation between (own) Grades and Neighbor Grades, is denoted by β_N, and the mean of the distribution of pseudosample coefficients, $E[\beta'_N]$. Since $E[\beta'_N]$ estimates the social correlation in the network without peer influence, then $\beta_N - E[\beta'_N]$ measures the peer influence in the network.

6.2.2 Shuffling Methodology

Shuffling has some potentially undesirable consequences, such as placing pseudosampled students in locations they would never visit in real life. Likewise, pseudosampled students have a more evenly distributed degree and average grade distributions (by the central limit theorem). Ideally, pseudosamples would randomize students' location (and social ties) without statistically interfering with relevant parameters (such as degree and grade distributions).

We control for this by restricting shuffle to students within building locations. So, for a given student, a pseudosample will never find her in at an access point she has never previously visited in real life; likewise, that student's usage in building B (classrooms) will all stay within that building to preserve the overall distribution of usage by location.

Restricted Shuffle

Randomization can change the structure of the network. For example, average usage per Access Point is likely to have larger variance in the real network than in pseudosamples. Great care must be taken to randomize the network in a manner that is orthogonal to the mode of influence. In this work, we restrict randomization of wifi sessions to the student, semester, and building of the session. In this way, no student will have a pseudosample session outside of where she might conceivably have actually gone in real life, but her relationships with other students will be broken up and shuffled in consequence of placing her at different Access Points, randomly throughout the buildings of her typical usage. Thus, we are not changing the nature of the students or the network (putting students from the library in the café), we are only changing the identity of the relationships. This allows us to argue that randomization removes peer influence but retains homophily.

6.2.3 Peer Definitions

Estimation of peer effects requires the definition of peers. Since this social network is inferred from the wifi usage data, we have no canonical "ground truth" as to who are friends with whom. The empirical nature of this work thus requires us to examine several definitions of peers, and then see how sensitive the results are to these definitions.

As a review from Chapter 3, we infer the social network of students using their co-wifi usage, and define a weighted edge in the graph as the number of 5-minute periods shared by those students. Every students' neighbors can be ordered by strength of relationship (cumulative time spent together). Some of these edges are certainly spurious, but it is expected that these edges would have a low weight. On the other hand, we may reasonably conclude that the strongest edge weights correspond to students' best "friends" (in terms of time spent physically together and using computers). The problem then becomes one of classification: given a student's distribution of strength of neighbors, which neighbors are true friends and which are noise?

Empirically, instead of trying to determine the correct threshold for "friendship", we simply repeat the analysis averaging neighbors' grades over different numbers of "friends".

6.2.4 Limitations

The Shuffle Test can be a powerful way to identify a causal effect in a social network. However, several limitations or caveats present themselves either methodologically or in consequence of our data.

Lagged

A key concern with the Shuffle Test is simultaneity: since own grades are jointly determined with neighbors' grades, it is difficult to determine the direction of causality. One way to overcome this is to regress own grades on neighbors' lagged grades.

Classrooms

Another concern unique to these data is the potential confounding of classroom wifi usage and true social influence. We are using wifi usage to proxy real human relationships, with the assumption that most relationships require physical proximity to be initiated and maintained. But with hundreds of students meeting for classes in the same building, it becomes problematic to assume that frequent time "together" attending lecture represents any relationship at all. It should be noted that Portuguese students, particularly in this time frame, are somewhat less likely to use laptops in class (if they bring them at all); nevertheless, this problem cannot be ignored.

6.3 Model

We examine several related models and methodological variations for robustness. Our canonical model measures peer performance as the average of the top three neighbors or "friends" (this is what is meant be 'means' in 'linear-in-means'), denoted Neighbor No. Courses and Neighbor Grades Points (and Neighbor Grade Points*), and estimates the causal peer effect under randomization. We also consider a lagged grades OLS model to control for simultaneity.

6.3.1 Linear-in-Means Peer Effect

This work employs the classic linear-in-means peer effect model (Sacerdote, 2011), defining Neighbor Grades as the average grade of students' neighbors in the network, for different numbers of neighbors (Best "Friend" to Top 5 "Friends"). Social correlation is captured by the coefficient of Neighbor Grades ($N_{i,t}$), β_N, in the linear model (Equation 4.1):

$$P_{i,t} = \beta_0 + N_{i,t}\beta_N + \boldsymbol{Z}'_{i,t}\boldsymbol{\beta_Z} + \epsilon_{i,t} \tag{6.1}$$

Equation 6.1 is a linear-in-means peer effect model commonly used in literature. We want to test if $\beta_N \neq 0$. Note that β_N itself is essentially a measure of social correlation, which says nothing about causality unless considered under randomization.

6.3.2 Nonlinear Peer Effects

More importantly, it would be interesting to know whether it is possible to design group assignment policies in ways that increase total productivity of the education production function. For instance, if by changing group assignments we can make some students learn more without adversely affecting other students. We look at heterogeneous effects by separating students by quantiles and comparing higher and lower performing groups.

The heterogeneous effects model is identical to 6.1, with only difference being in the observations which are included for regression, with the lower and upper quantiles being selected from the data prior to regression. Quantiles are taken on Application Score, a measure of prior (high school) performance, to avoid simultaneity in own and neighbor grades being jointly determined. Note that the Neighbor Grade covariate is only calculated once for the data set and for each pseudosample. This necessitates careful interpretation of the resulting peer effect, specifically, that a positive peer effect is *good* for "bad" students and *bad* for "good" students.

As will be discussed in Chapter 7, we look at the heterogeneous effect by splitting the population into upper and lower quantiles using quantiles 0.5, 0.33, and 0.25 as cutoffs.

6.3.3 Peer Effects with Lagged Grades

The lagged grades model is almost identical to the linear-in-means model in Equation 6.1, except that Lagged Neighbor Grades is computed from the prior semester:

$$P_{i,t} = \beta_0 + N_{i,t-1}\beta_N + \boldsymbol{Z}'_{i,t}\boldsymbol{\beta_Z} + \epsilon_{i,t} \tag{6.2}$$

Lagging Neighbor Grades avoids endogeneity by which own and neighbor grades are jointly determined. Since Lagging and Randomization are two methods to accomplish the same goal, it is not appropriate to apply the Shuffle Test to Equation 6.2; however, results from both methods will be discussed in Chapter 7.

6.3.4 Peer Effects without Classroom Usage

6.4 Empirical Strategy

Following the generation of data as discussed in Section 6.2 and Chapter 3, the empirical strategy in this part is to extract the student cross-section for each pseudosample, and run the models in Section 6.3. We then collect the resulting distribution of coefficients of neighbors' grades and plot the results in various ways as will be shown in Chapter 7.

As we saw in Chapter 4, the data provide several different measures for our parameters, so, naturally, we will want to examine each of these variations for sensitivity and robustness.

Chapter 7

Peer Effects: Results

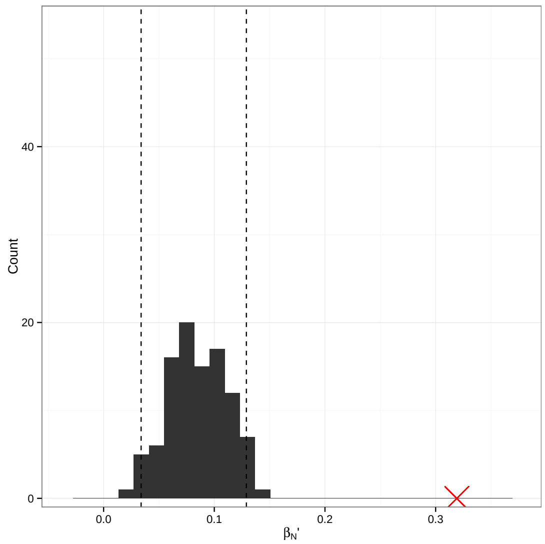
7.1 Overview

In this chapter we examine the empirical results stemming from the discussion in Chapter 6. We proceed in two parts, first examining the results with randomization, and then comparing these results with results using lagged Neighbor Grades.

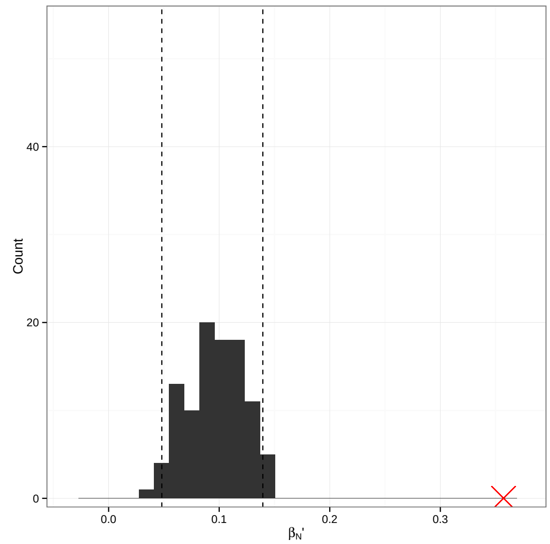
7.2 Results with Randomization

Figure 7.1 illustrates randomization and the use of the Shuffle Test to estimate the peer effect in the social network for Fall 2008. First, β_N is estimated using Equation 4.1, and is plotted as $\times$. Then, the same model is used 100 times to estimate β'_N. This is plotted as a histogram in the same figure. A 95% confidence interval is plotted over this histogram as dotted vertical lines. This is done for three different measures of performance, as noted in the sub-figure captions. The peer effect is estimated to be the difference between the true value and the distribution. This yields a mean peer effect[1] and confidence interval, which

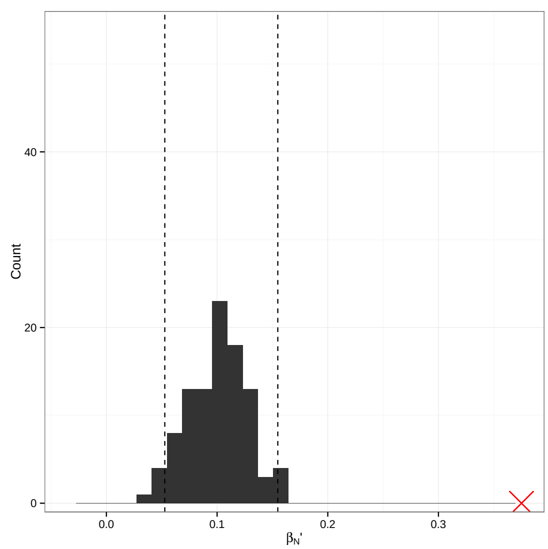
[1](Belo et al., 2014) show in other work that under general conditions this is not a mean effect, but rather a lower bound.



(a) No. Courses



(b) Grade Points



(c) Grade Points*

Figure 7.1: Real and Randomized β_N' for Fall 2008 and Top 3 Neighbors. True β_N denoted by $\times$.

will be shown as a point with error bars in forthcoming plots.

7.2.1 Linear-in-Means Peer Effect

We first examine the average peer effect over all students in order to test our methodology, verify the presence of a peer effect that we would expect in theory, and estimate the magnitude of that effect. Our methodology also allows us to test the sensitivity of our parameters.

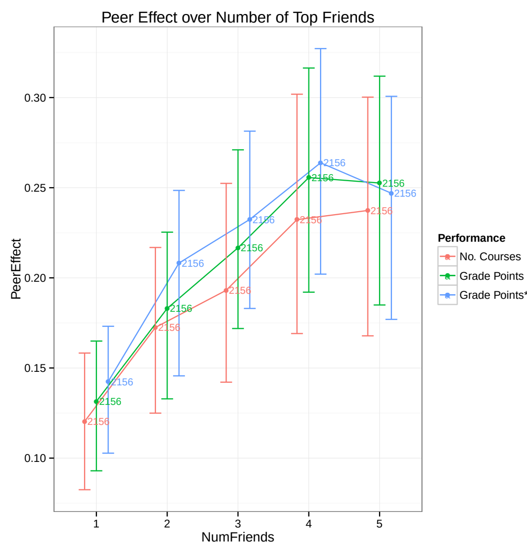


Figure 7.2: Peer Effect by Performance over No. Neighbors.

In Figure 7.2, we estimate the peer effect for all students and for all measures of performance across five different thresholds for "friendship". We note a statistically significant, positive peer effect for all measures ranging from 0.15-0.40, giving us confidence that a measurable, non-zero effect does exist. We see some sensitivity to the definition of "friendship"; results based only on the "best" friend's grades show the smallest peer effect, and this increases with an increasing number of "friends". Finally, we generally see the smallest effect measuring by No. Courses, but note that all three measures of performance are statistically indistinguishable.

7.2.2 Heterogeneous Effects

Peer effects are predicted by theory and experienced in almost every domain, and we have merely verified that fact in the prior section. What makes this work particular interesting is the present effort to differentiate the peer effect for different types of students. This differentiated peer effect can lead to very interesting policy implications. As discussed in Section 6.3.2, we now turn to the analysis and discussion of heterogeneous peer effects. For brevity, we present and discuss only results on No. Courses, since this appears to be the most modest (smallest) estimator, and we have not seen any statistically difference between the measures of performance. Table 7.1 summarizes the results of heterogeneous randomizations.

Table 7.1: Nonlinear Peer Effect Results

Curr. Year	Quantile	Low/High	All Data			No Classroom		
			No. Courses	Grade Points	Grade Points*	No. Courses	Grade Points	Grade Points*
Half	1-2	L	0	0	0	0	0	0
Half	1-2	H	+	+	+	0	0	0
Half	3-4	L	++	++	++	+	+	+
Half	3-4	H	0	+	+	0	0	0
Tercile	1-2	L	0	0	0	0	0	0
Tercile	1-2	H	+	+	+	+	0	0
Tercile	3-4	L	++	++	++	+	+	+
Tercile	3-4	H	+	+	+	0	0	0
Quartile	1-2	L	0	0	0	0	0	0
Quartile	1-2	H	+	+	+	0	0	0
Quartile	3-4	L	++	++	++	+	+	+
Quartile	3-4	H	+	+	+	0	0	0

Halves

We start the discussion of heterogeneous effects by splitting the data under observation into upper and lower halves in the Application Score distribution. The first half corresponds to below-median ("bad") students as measured by their high school performance and college application scores (Application Score); the second half corresponds to above-median ("good") students.

Figure 7.3 shows the peer effect over No. Neighbors, split across lower and upperclassmen. We find that among lowerclassmen, above-median students have a modestly higher peer ef-

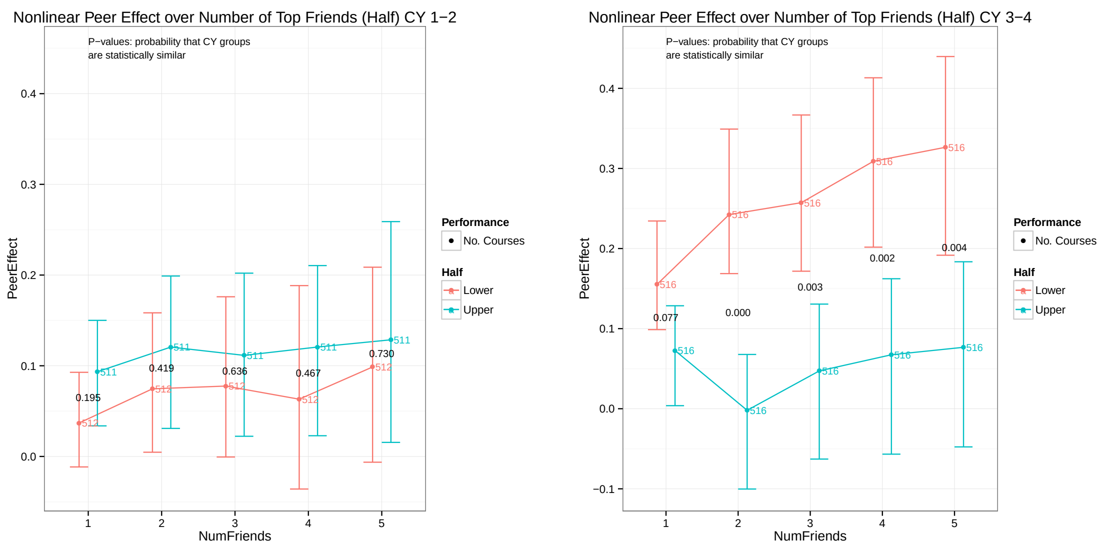


Figure 7.3: No. Courses Peer Effect by Halves over No. Neighbors, split by Curricular Year.

fect, but the difference is not statistically significant (p-values comparing the two halves are shown in black). Note that above-median students have a positive, statistically significant peer effect, whereas below-median performers are statistically zero. However, among upperclassmen we see a persistent and statistically significant difference across across all No. Neighbors, with below-median students having a statistically larger peer effect (whereas above-median students being statistically zero). These results clearly suggest that tracking students is advantageous in early curricular years, whereas grouping is advantageous in later curricular years.

Terciles

Observing that below-median students exhibit greater peer effects in later curricular years, we now investigate this phenomenon more granularly by looking at other quantile divisions of students. In this section we examine the tercile results. In this analysis we observed that the lowest tercile exhibited the same strongly positive peer effect, so we decided to group the second and third terciles for clarity and simplicity. Thus, Figure 7.4 shows the peer effect by No. Courses over Curr. Years, divided into lower tercile and upper two terciles. This result is similar to Figure 7.3. Note that in the lower curricular years, tercile divisions are not statistically different, although the “good” students show a non-zero peer effect. In later curricular years we see a clear division in the peer effect, with “bad” students having a large, positive, statistically significant effect, that is also statistically different from the “good” students. More importantly, this suggests a policy of grouping students in later curricular years. Note that unlike Figure 7.3, “good” upperclassman also have a modestly significant positive peer effect. This most likely comes from including below-median observations in the “good” group, and illustrates that the large peer effect is not entirely driven by outliers.

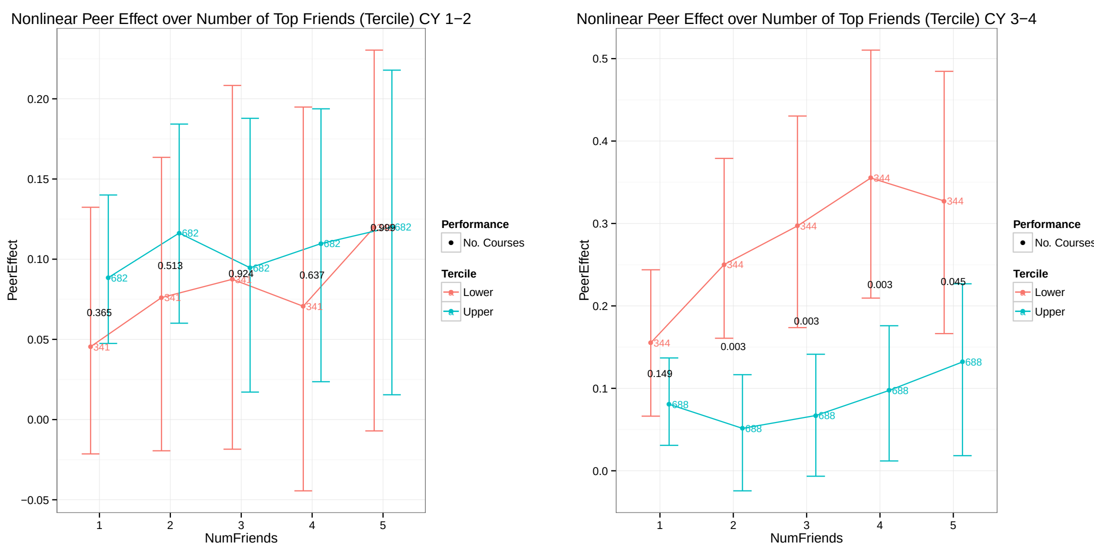


Figure 7.4: No. Courses Peer Effect by Terciles over No. Neighbors, split by Curricular Year.

Quartiles

We conclude this work on heterogeneous effects by trying to further isolate what portion of students exhibit a positive peer effect by looking at quartiles. We again show the lower quartile and combine the other three quartiles for clarity. Figure 7.5 shows the quantile effect. We see a similar pattern emerging, with “bad” lowerclassmen showing a statistically zero peer effect, and "good" lowerclassmen showing a positive, statistically significant peer effect of 0.11-0.13. This provides support for tracking in early curricular years, but again, the two groups are not statistically different.

For upperclassmen, Figure 7.5 shows a strongly positive and significant peer effect for "bad" students" and a strong statistical difference from the "good" students. Interestingly, we also see that the "good students" (now including the second quartile below the median) shows a modestly positive, but statistically significant, peer effect. This stems in part from including observations below the median that show a positive peer effect, but note that the "bad" student peer effect is not diluted at all, and is in fact the largest among the three quartile divisions for upperclassmen. These results further support a policy of tracking in early curricular years, and grouping in later curricular years.

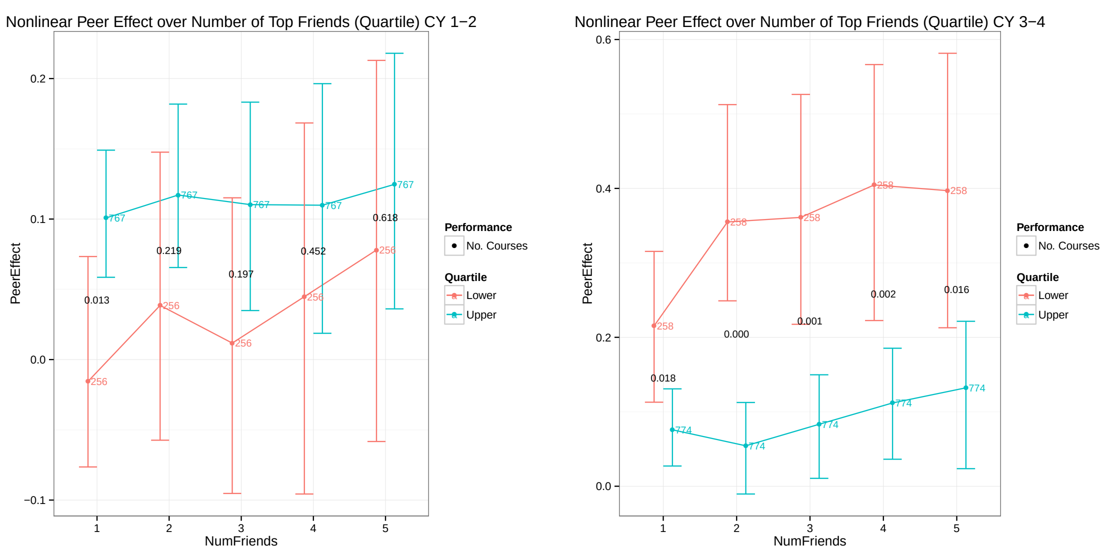


Figure 7.5: No. Courses Peer Effect by Quartiles over No. Neighbors, split by Curricular Year.

7.2.3 No Classroom

As discussed in Section 6.2.4, a key concern in this work is that we are detecting spurious in-class usage as "friendship" when in fact there may be no peer effect at all. In this section we investigate this issue by removing classroom usage from the wifi data before creating the social network (and before randomization).

Figure 7.6 shows the peer effect for all students across different No. Neighbors. Although smaller in magnitude than Figure 7.2, we see statistical significance for all but one case. Recall that Grade Points is an upper bound and Grade Points* is a lower bound, so even in the 5 Neighbors case there likely to be a true effect. Even so, the various measures of performance are not statistically different.

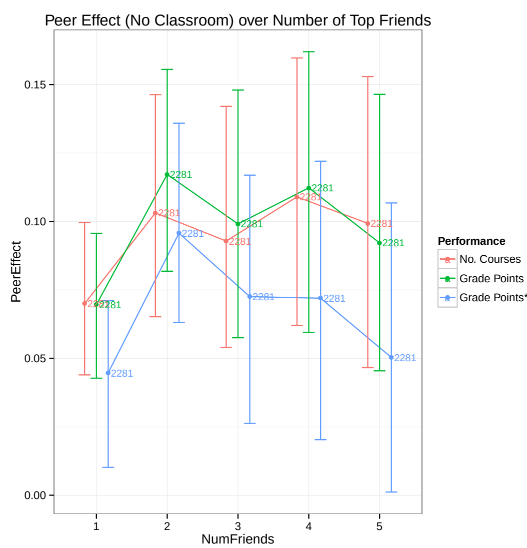


Figure 7.6: Peer Effect without Classroom by Performance over No. Neighbors.

Halves

For robustness, we check out the various quantile work as before, repeated on that subset of the data which excludes in-class usage. Note that removing classrooms from the data is an aggressive way to control for spurious relationships stemming from in-class usage, in that it removes all classroom usage and some non-classroom usage. It also reduces the strength of true friendships and reduces their statistical power. So in Figure 7.7, with only the exception of the "bad" best friend (which is statistically positive), underclassmen show no peer effect outside of the classroom. This would suggest that much of underclassmens' interaction happens through their in-class cohorts. Results for upperclassmen are also weaker, but not insignificant. It appears that when removing in-class usage from the analysis, upperclassmen show a statistically positive and statistically different peer effect for "bad" students over "good" for the top 2 or 3 "friends". This would suggest that students have slightly fewer "friends" (from FEUP) outside of their classes; that is, that attending class expands their pool of "friends" somewhat.

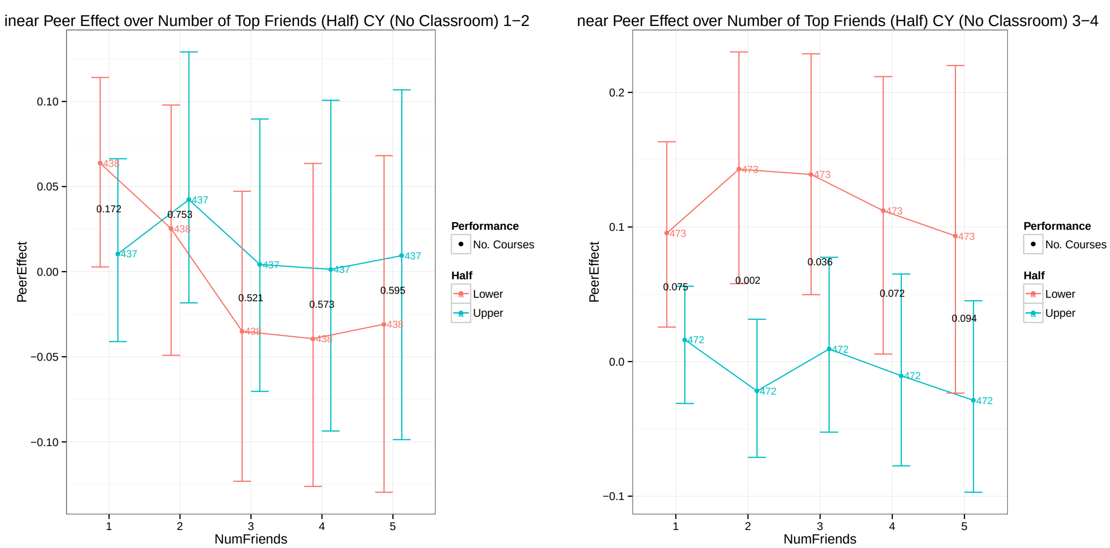


Figure 7.7: No. Courses Peer Effect by Halves over No. Neighbors, split by Curricular Year (No Classroom).

7.3 Results with Ordinary Least Squares and Lagged Variables

Our basic goal is to identify a causal peer effect using randomization, but as noted, randomization has some potential side effects, and care must be taken to run appropriate robustness checks to verify that reported results are real. In this section we seek to control for simultaneity by regressing student grades on the prior semester grades of their current neighbors. The idea here is that for whatever reason students are now connected to each other, their grades are not jointly determined. Note that it does not make so much sense to apply randomization and lags at the same time, since they both accomplish essentially the same goal (reduce endogeneity) and can lead to a loss of statistical power that is not attributable to the lack of an effect. In this section we examine the linear-in-means peer effect through OLS regression on lagged grades.

7.3.1 Linear-in-Means Peer Effect

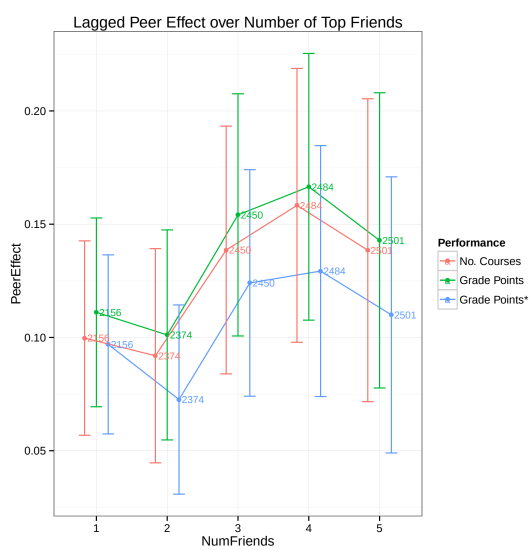


Figure 7.8: Lagged Peer Effect by Performance over No. Neighbors.

In Figure 7.8, we find a statistically significant, positive peer effect for all "friends" and across all measures of performance. Although we see some fluctuation across No. Neighbors, the difference is not very statistically significant. We see again that peer effect is estimated most strongly by No. Courses, with Grade Points serving as an upper bound and Grade Points* as a lower bound. These results are consistent with Figures 7.2 and 7.6.

7.3.2 Summary

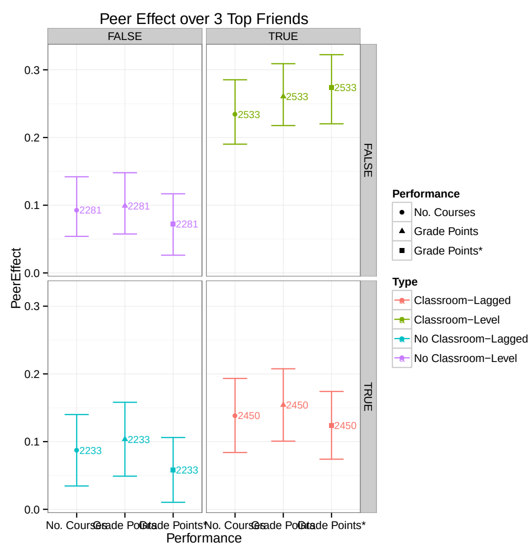


Figure 7.9: Randomized and Lagged Peer Effect.

Figure 7.9 summarizes all of the homogeneous peer effect results and robustness checks for the No. Neighbors = 3 case, and shows that although the estimated magnitude of peer effect varies by method, all robustness checks show a positive, statistically significant result.

7.4 Discussion

We find strong evidence of a causal peer effect using randomization. This result verifies the presence of a peer effect and the validity of the identification strategy. As expected from the literature, heterogeneous models also provided interesting results (Calvó-Armengol, Patacchini, & Zenou, 2009; Burke & Sass, 2013; Jain & Langer, 2013).

One clear next step is to repeat the analysis across semesters of available data. Another important next step is to examine the time dimension using dynamic social network analysis techniques (Aral et al., 2009; La Fond & Neville, 2010); that is, by creating a panel of student-semesters and repeating the analysis in Chapter 5, controlling for neighbors' grades instead of wifi.

Policy Implications of Peer Effects Results

This work shows clear evidence of a peer effect among FEUP students. We seek to differentiate this effect by understanding which portions of the population manifest it most strongly. Recall that the linear-in-means peer effect is a measure of correlation with the mean, so a "positive" peer effect simply means that "bad" students are helped (pulled up toward the mean) and "good" students are hurt (pulled down to the mean).

We find that underclassmen show a smaller, often negligible peer effect, and it is usually stronger for the "good" students. This provides modest support for a policy of tracking—placing the better students with better students in early courses. This could be done, for instance, through an honors freshman project course, in which the best applicants qualify for participation.

We also find that among upperclassmen, poor students consistently show a strong peer effect, and good students sometimes show a modest effect. This suggests a clear policy of non-tracking, or grouping, in which all upperclassmen are pooled together so that bad students are helped in their work without overly harming the good students.

Chapter 8

Conclusion & Policy Implications

8.1 Conclusion

In this work we examine ICT and peer effects in the context of higher education. We use a large data set that is relatively uncommon in higher education literature, yet this research is replicable using similar data sets. This thesis also demonstrates the practicality of doing policy research on individuals while maintaining anonymity, and may be used as a case study to encourage the release of similar data sets at other institutions.

This thesis is more methodologically rich than many prior works, particularly on the ICT side. This is due in large part to the paucity of large-scale panel data in higher education. Econometric methods as employed in this thesis allow us to argue for a causal ICT and peer effect. In addition, they allow us to approach a policy problem at a larger scale and from a policy standpoint—at the level of the institution. On the other hand, the econometric approach leaves us with somewhat to be desired in answering theoretical questions as to why we obtain the results we do, and so as in all things the quantitative approach can be balanced with qualitative insight (such as by open surveys and interviews).

In Chapter 5, we demonstrate a positive, statistically significant relationship between wifi usage and student performance at the university level. In Chapter 7, we show a positive, significant, and causal peer effect between students.

Using both OLS and PSM, we demonstrate that wifi usage has a positive relationship with academic performance among FEUP students. This effect is generally larger for Hours online than Megabytes transferred, as also for Grade Points over No. Courses. We see a larger effect for daytime users over nighttime users, for more mature, higher curricular year students, and with substantial variation across different academic majors measuring in terms of Hours. An ordering of wifi productivity among majors can be discerned, with the more computer-oriented majors showing greater productivity online.

Of central interest in economic literature is the estimation of causal effects. The OLS-FD model is effective in differentiating the wifi effect, and the use of First Differences allows us to claim more than a mere correlation, but we likewise cannot argue that these models are completely free from endogeneity. First Differences controls for student time-constant effects (nature, aptitude, intelligence), but both past (and present) grades and wifi usage are jointly determined, and we cannot say whether an increase (decrease) in wifi usage *causes* an increase (decrease) in grades, or vice versa, but we can say that such changes are not due to intrinsic student characteristics.

Using Propensity Score Matching, we find corroborating evidence of the OLS-FD results, and also show that Laptop ownership has a statistically zero (or nearly zero) relationship with grades. We find that wifi usage (and not mere laptop ownership) drives the observed positive effect.

We infer a proxy social network from the true student social network through co-wifi usage among students, using the number of common 5-minute periods at the same physical Access Point as the strength of relationship (edge weight). We use this graph to calculate the average of neighbor attributes (Grades and Application Score) for various types of relationships. We believe this is a novel method for inferring a social network from wifi data.

We find strong evidence of a causal peer effect using randomization over linear-in-means models. We find a positive and statistically significant effect for all measures of performance. Great care is taken to apply randomization in a way that does not create misleading results, such as significantly changing the composition of students who connect to a given Access Point.

This work employs linear, heterogeneous models to determine the peer effect on different types of students. Further, we subdivide observations to analyze the peer effect by Curr. Year over upper and lower ("good" and "bad") quantile students. We find a small, positive, and statistically significant peer effect for upper quantile underclassmen, providing support for a policy of tracking in the earlier curricular years. We find a positive, statistically significant peer effect for lower quantile upperclassmen, providing support for a policy of grouping the later curricular year students.

This study is unique in its breadth, being one of relatively few studies using an individualized, campus-wide panel of ICT and grades, Marmaros and Sacerdote (2006) is the only other example we know of. We expect this work will contribute to higher education policy, technology policy, and policy to clarify the utility of wifi as an increasingly ubiquitous educational technology.

8.2 Policy Implications

In discussing policy implications it is essential to understand what the policy goal actually is. Is it to increase the average grades of all students? Or is it to increase the number of students passing? In either case, several clear policy recommendations stem from this work, as pertaining to both ICT effects and peer effects.

On the ICT side, this work validates the widespread deployment of wifi in higher education, both in terms of the widely understood benefits of ICT literacy and democratized access to information, and also in terms of actual academic performance—an issue which has been widely debated.

Other policy implications stem from this work on differentiated wifi effects, specifically that daytime use is more productive than nighttime use, and upper curricular years are more productive than lower. While throttled nighttime access is unlikely to be a successful policy, there is potential for a first or second curricular year internet resource and/or time management training to help students use ICT resources more productively and successfully. PSM results suggest that wifi usage is more important than laptop ownership. This would suggest a program of laptop lending.

On the peer effects side, this work clearly demonstrates the importance and relevance of peer effects in higher education. This affects many aspects of policy, from group work to tracking. Heterogeneous effects show that there is some potential benefit for tracking earlier curricular year students, since the better students receive a larger peer effect than the worse students. We also find evidence that it is beneficial to group students in later curricular years, since "bad" students have a much larger peer effect than "good" students, and therefore stand to benefit more from grouping than they would hurt those good students. To our knowledge, this is the only work that recommends a heterogeneous policy on tracking and grouping.

8.3 Further Research

This work identifies several avenues of additional research. Unfortunately, in some cases this research cannot be done with these data.

First, we do not observe internet usage over the fixed (Ethernet) network; however, we believe that controlling for major allows us to capture some of this effect because students in the same major are likely to use the fixed network in similar ways, e.g. Electrical Engineering and Computer Science use alike and more to complete programming and networking related courses, other major are likely to use less. Further, a student survey completed in 2012 leads us to believe that students are most likely to use a laptop when they use a computer on campus, and that they are most likely to connect to wifi when they use a laptop. While more complete data is always desirable in any project, we feel we have used the available data well.

Second, for obvious reasons, we do not observe wifi usage for students who do not use wifi; therefore, we are not able to say anything about those who do not use wifi. If there are characteristic reasons why students do not use wifi, we would want to identify and account for them. This could entail the use of a Heckman selection model to evaluate the probability of being a wifi user, prior to observing wifi.

Third, the data do not contain so robust covariates as would be desirable. If we could, we would be interested to examine the wifi effect across other socioeconomic factors such as gender and parental education or income.

Fourth, the data, unfortunately, provide no information, besides session duration and bytes transferred, on what students do with the wifi they use. If we could we would like to look at the specific applications, websites, and services students use, and evaluate more granularly what has a positive and negative impact on performance.

Finally, social network analysis was done individually at the semester level. An important next step is to examine the time dimension using dynamic social network analysis techniques. For example, we can see how (change of) group membership and friendships affects grades over time. We can examine the persistence of a prior effect. We can look at how (dynamic) network structure affects a nodes influence. Some of what we learned from the work on peer effects can be applied to the earlier work on ICT effects (such as testing effects on both Grade Points and Grade Points*).

Appendix A

ICT Effects: Additional Results

A.1 Time on Campus

Table A.1: Time on Campus

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	ΔNo. Courses	ΔNo. Courses	ΔNo. Courses	ΔNo. Courses	ΔGrade Points	ΔGrade Points	ΔGrade Points	ΔGrade Points	ΔGrade Points*	ΔGrade Points*	ΔGrade Points*	ΔGrade Points*
ΔHours	0.104***	0.131***			0.113***	0.140***			0.0917***	0.121***		
	(0.0160)	(0.0174)			(0.0166)	(0.0180)			(0.0156)	(0.0187)		
ΔMegabytes			0.0181	0.0150			0.0130	0.00586			0.0147	0.00992
			(0.0136)	(0.0155)			(0.0139)	(0.0156)			(0.0131)	(0.0170)
ΔCampus Hours	0.0549***	0.0442***	0.0985***	0.100***	0.0556***	0.0493***	0.105***	0.111***	0.0472***	0.0511***	0.0860***	0.103***
	(0.0149)	(0.0164)	(0.0142)	(0.0156)	(0.0149)	(0.0161)	(0.0143)	(0.0154)	(0.0141)	(0.0156)	(0.0134)	(0.0151)
Application Score		-0.0331***		-0.0354***		-0.0479***		-0.0506***		-0.0582***		-0.0604***
		(0.0112)		(0.0111)		(0.0116)		(0.0114)		(0.0123)		(0.0122)
Constant	-0.160***	0.200	-0.140***	0.184	-0.132***	0.153	-0.109***	0.138	-0.117***	0.0797	-0.0995***	0.0657
	(0.0119)	(0.142)	(0.0112)	(0.141)	(0.0119)	(0.140)	(0.0112)	(0.138)	(0.0121)	(0.131)	(0.0115)	(0.129)
Cohort Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Curricular Year Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Major Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Semester Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N (student-semesters)	6399	5313	6399	5313	6399	5313	6399	5313	6399	5313	6399	5313
N (students)	3021	2413	3021	2413	3021	2413	3021	2413	3021	2413	3021	2413

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2 Usage Ratio (Bandwidth)

Table A.2: Usage Ratios (Bandwidth) (Quadratic)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ΔNo. Courses	ΔNo. Courses	ΔGrade Points	ΔGrade Points	ΔNo. Courses	ΔNo. Courses	ΔGrade Points	ΔGrade Points
ΔMB/Hours Ratio	0.0379*	0.0357*	0.0407*	0.0381*	0.0378*	0.0355*	0.0405*	0.0379*
	(0.0213)	(0.0185)	(0.0227)	(0.0199)	(0.0213)	(0.0185)	(0.0227)	(0.0199)
Application Score		-0.0332***		-0.0483***		-0.0331***		-0.0481***
		(0.0111)		(0.0114)		(0.0111)		(0.0114)
ΔMB/Hours Ratio (Quadratic)					-0.00165	-0.00692	-0.00764	-0.0107
					(0.0142)	(0.0147)	(0.0148)	(0.0146)
Constant	-0.133***	0.128	-0.102***	0.0811	-0.133***	0.129	-0.101***	0.0832
	(0.0111)	(0.141)	(0.0111)	(0.138)	(0.0113)	(0.141)	(0.0114)	(0.138)
Cohort Dummies	No	Yes	No	Yes	No	Yes	No	Yes
Curricular Year Dummies	No	Yes	No	Yes	No	Yes	No	Yes
Major Dummies	No	Yes	No	Yes	No	Yes	No	Yes
Semester Dummies	No	Yes	No	Yes	No	Yes	No	Yes
N (student-semesters)	6425	5337	6425	5337	6425	5337	6425	5337
N (students)	3030	2421	3030	2421	3030	2421	3030	2421

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

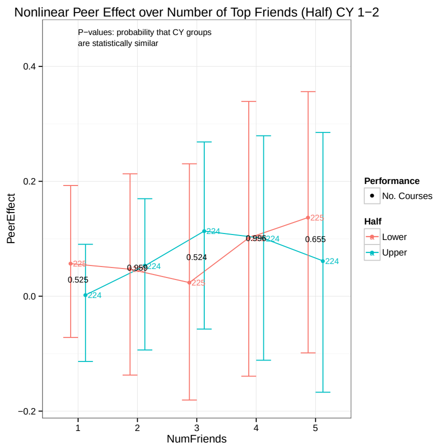
Appendix B

Peer Effects: Additional Nonlinear Results

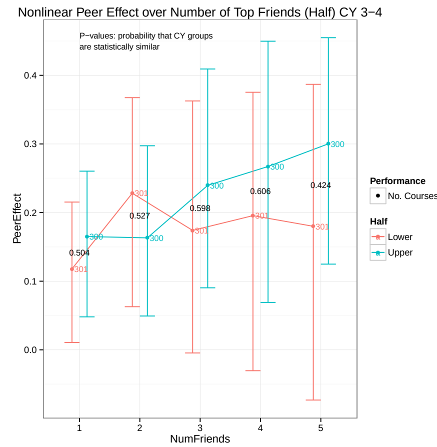
The main text shows nonlinear effects only on No. Courses. In this appendix, we show qualitatively similar results on Grade Points and Grade Points*.

B.1 Nonlinear Results with All Wifi Usage (Fall 2006)

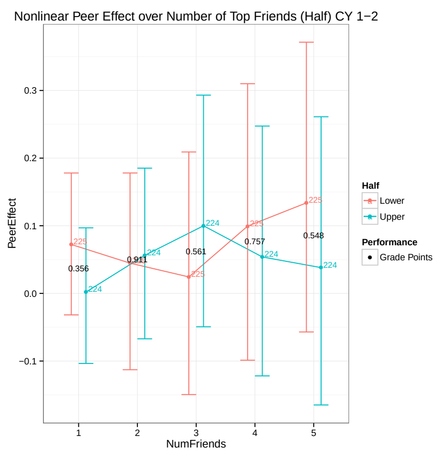
B.1.1 Halves



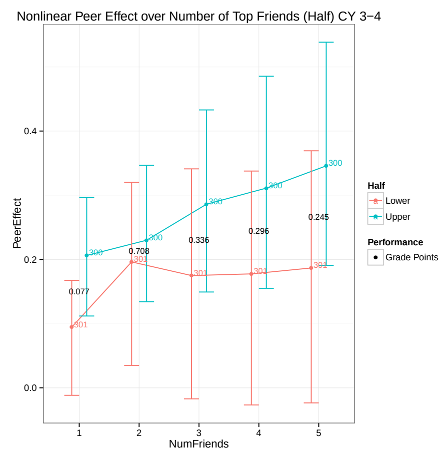
(a) No. Courses



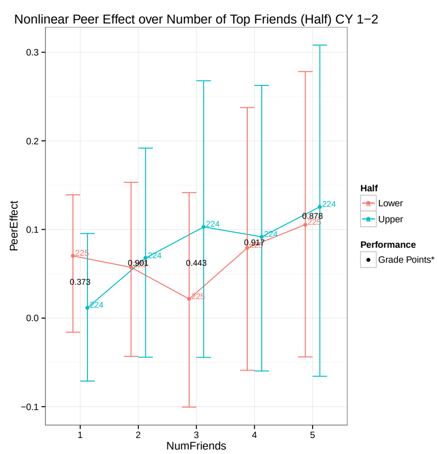
(b) No. Courses



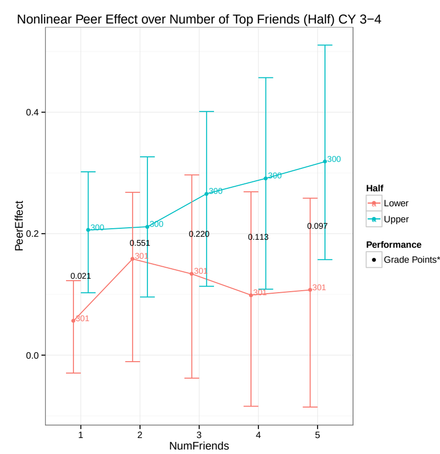
(c) Grade Points



(d) Grade Points



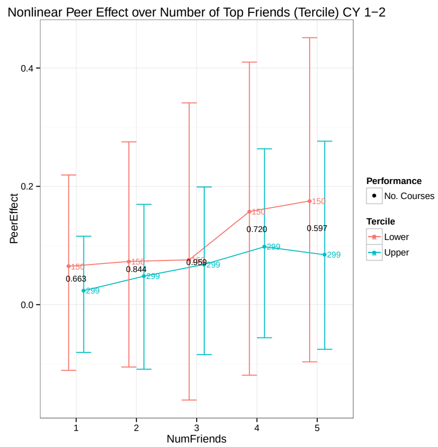
(e) Grade Points*



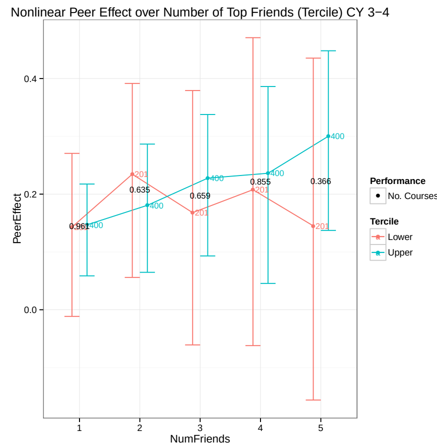
(f) Grade Points*

Figure B.1: Peer Effect by Halves over No. Neighbors, split by Curricular Year.

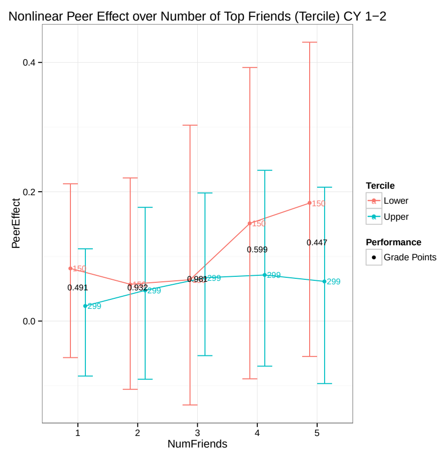
B.1.2 Terciles



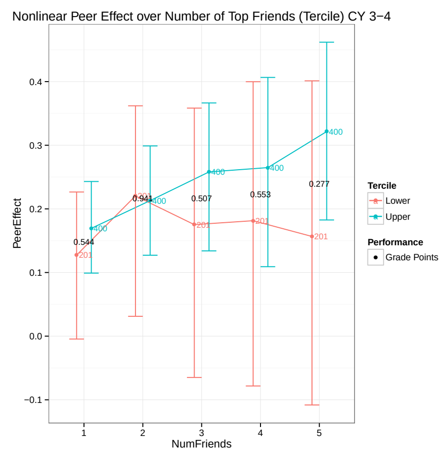
(a) No. Courses



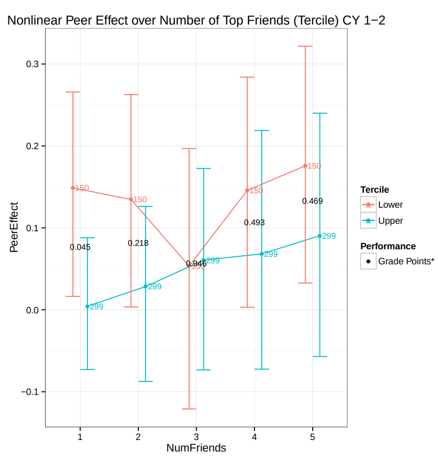
(b) No. Courses



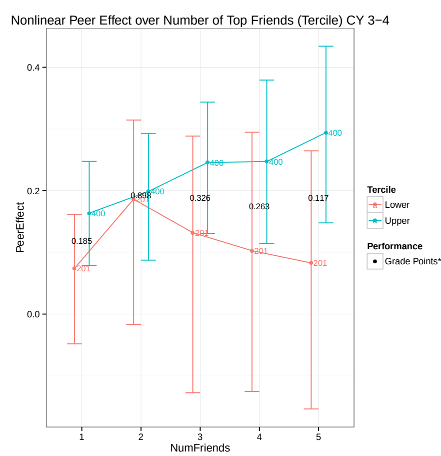
(c) Grade Points



(d) Grade Points



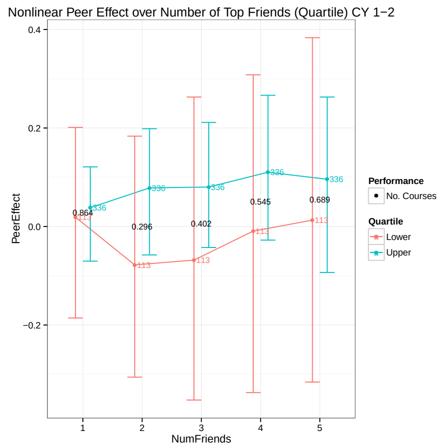
(e) Grade Points*



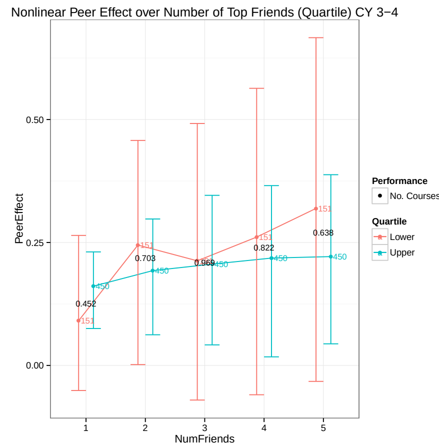
(f) Grade Points*

Figure B.2: Peer Effect by Terciles over No. Neighbors, split by Curricular Year.

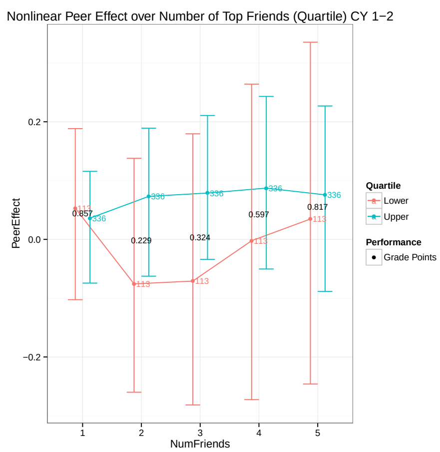
B.1.3 Quartiles



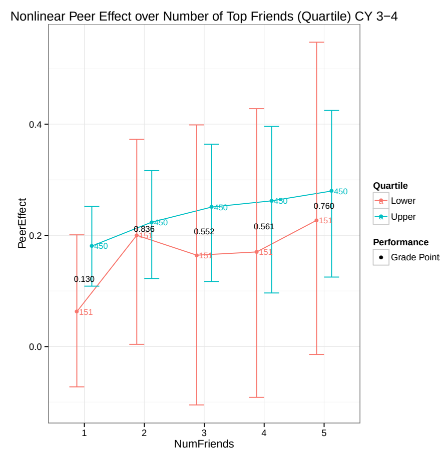
(a) No. Courses



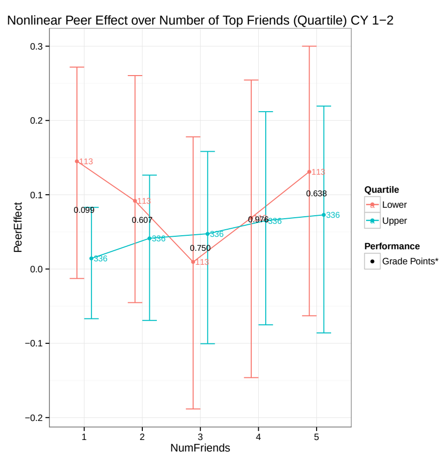
(b) No. Courses



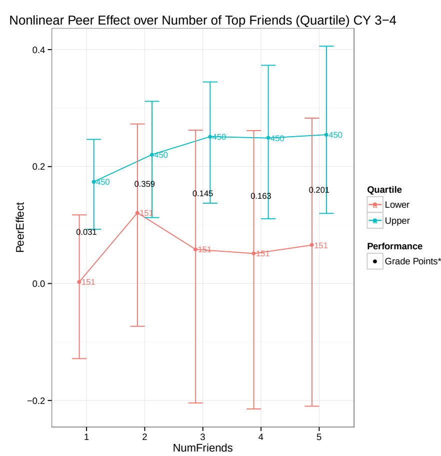
(c) Grade Points



(d) Grade Points



(e) Grade Points*

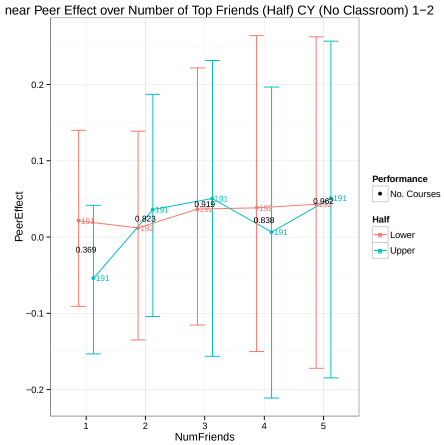


(f) Grade Points*

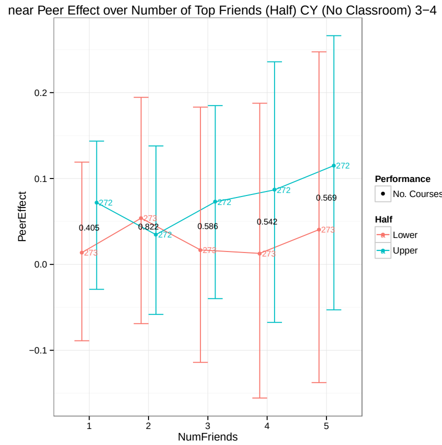
Figure B.3: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year.

B.2 Nonlinear Results without Classroom Usage (Fall 2006)

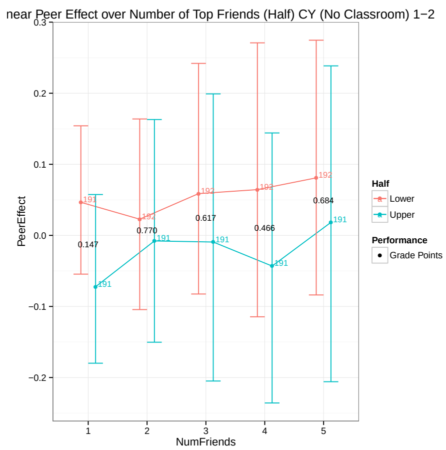
B.2.1 Halves



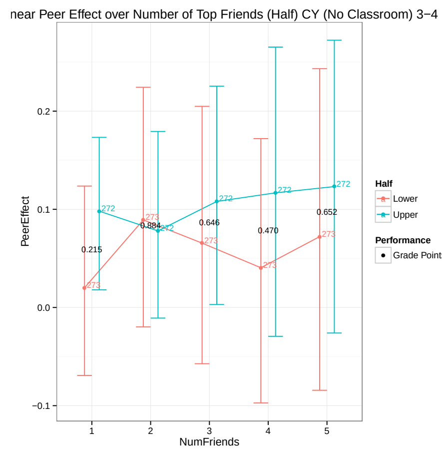
(a) No. Courses



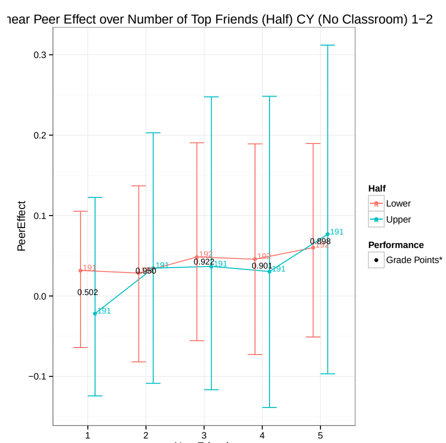
(b) No. Courses



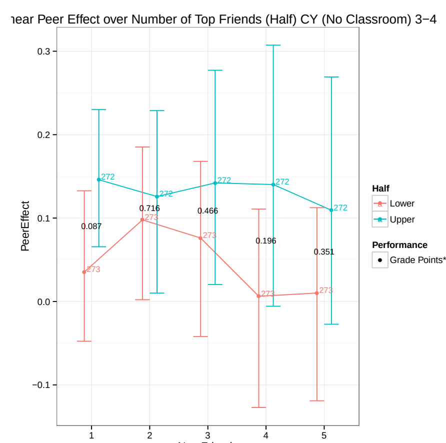
(c) Grade Points



(d) Grade Points



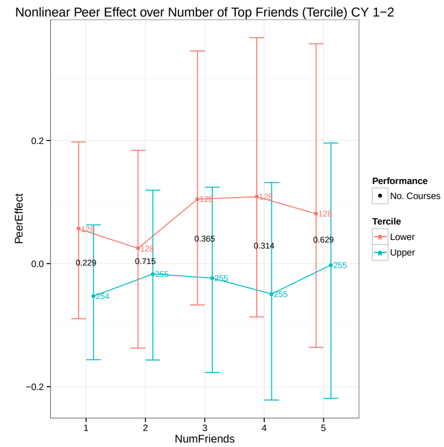
(e) Grade Points*



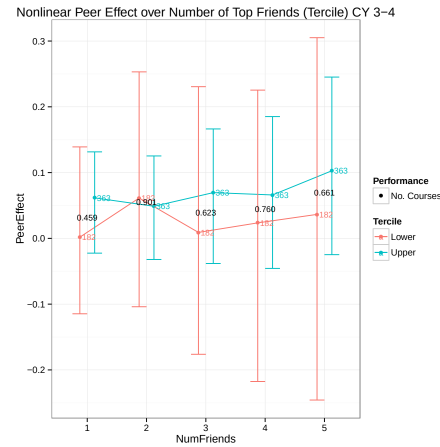
(f) Grade Points*

Figure B.4: Peer Effect by Halves over No. Neighbors, split by Curricular Year (No Classroom).

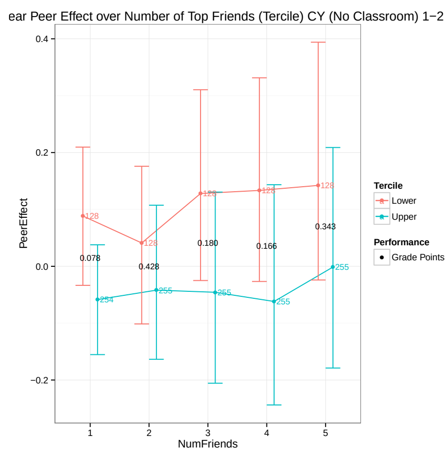
B.2.2 Terciles



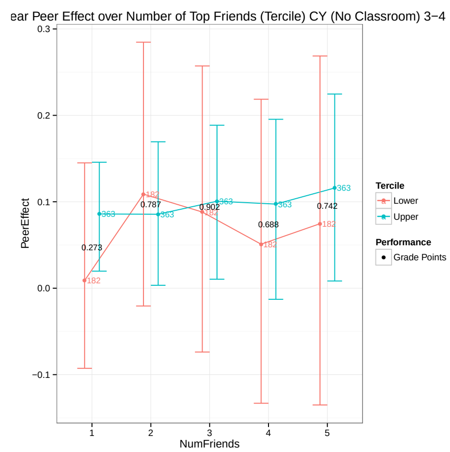
(a) No. Courses



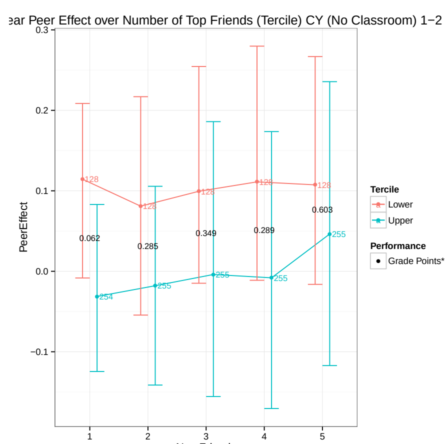
(b) No. Courses



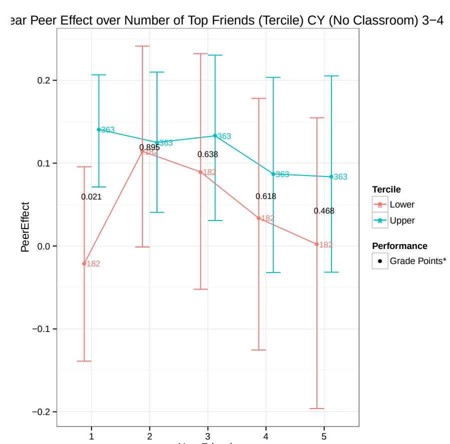
(c) Grade Points



(d) Grade Points



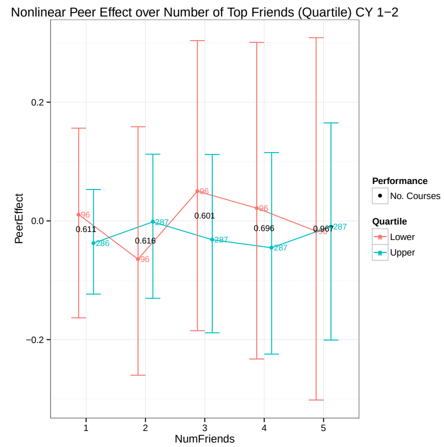
(e) Grade Points*



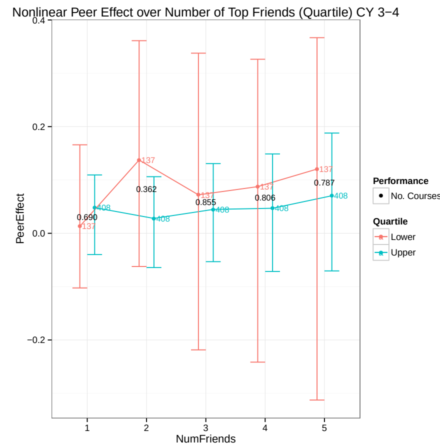
(f) Grade Points*

Figure B.5: Peer Effect by Terciles over No. Neighbors, split by Curricular Year (No Classroom).

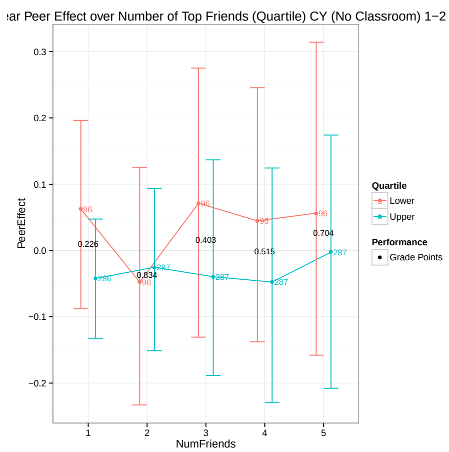
B.2.3 Quartiles



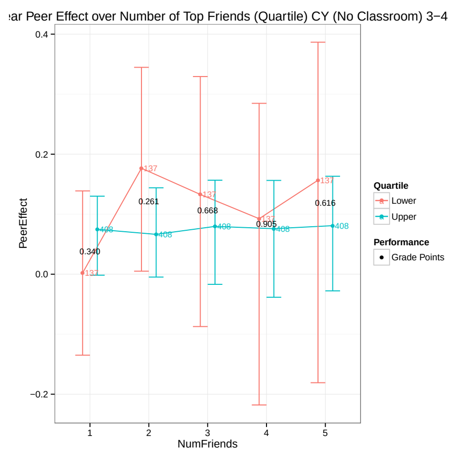
(a) No. Courses



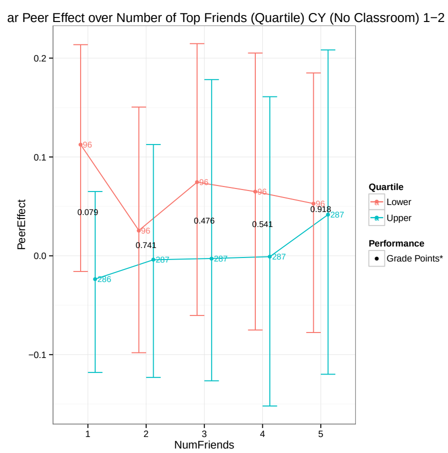
(b) No. Courses



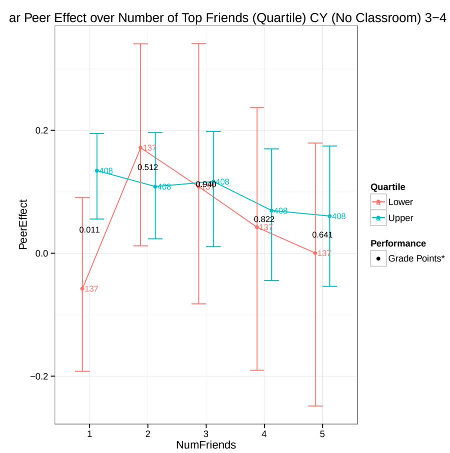
(c) Grade Points



(d) Grade Points



(e) Grade Points*

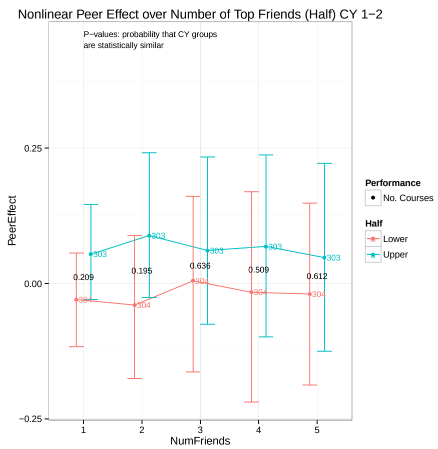


(f) Grade Points*

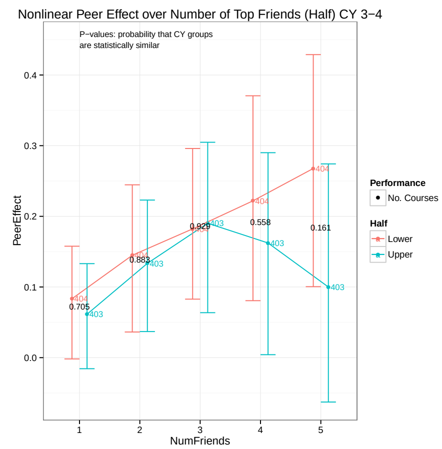
Figure B.6: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year (No Classroom).

B.3 Nonlinear Results with All Wifi Usage (Spring 2007)

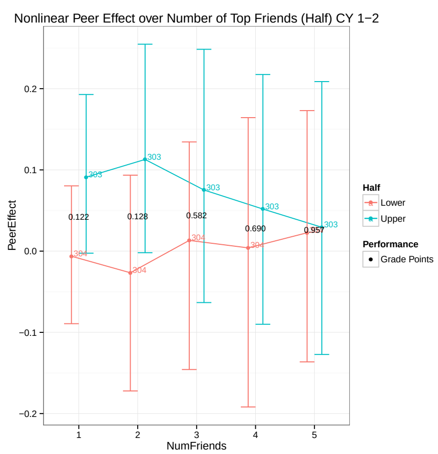
B.3.1 Halves



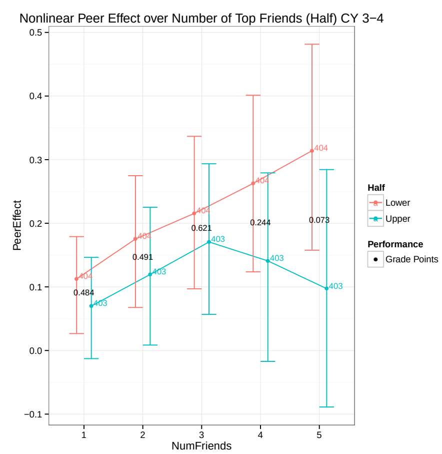
(a) No. Courses



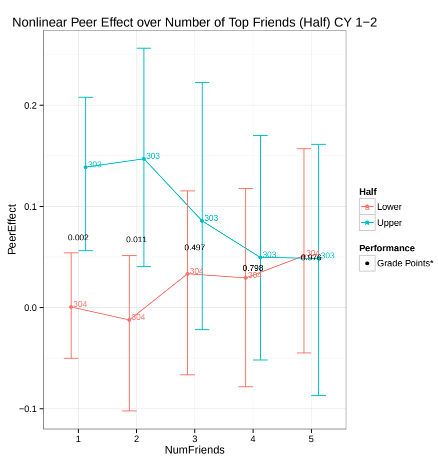
(b) No. Courses



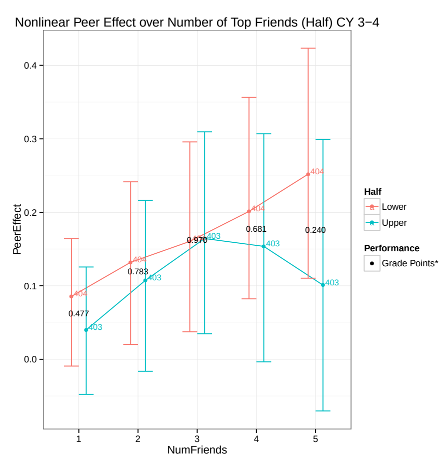
(c) Grade Points



(d) Grade Points



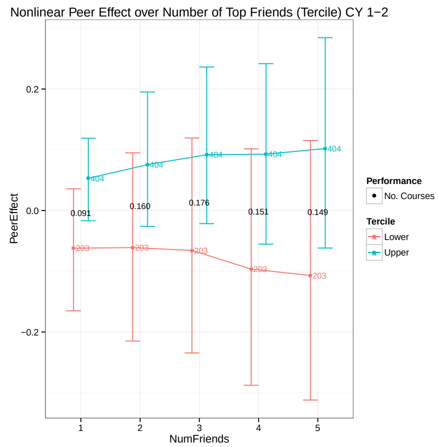
(e) Grade Points*



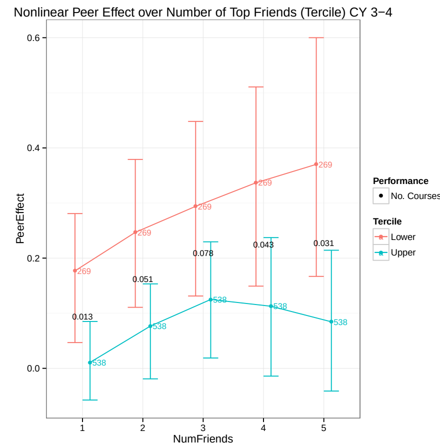
(f) Grade Points*

Figure B.7: Peer Effect by Halves over No. Neighbors, split by Curricular Year.

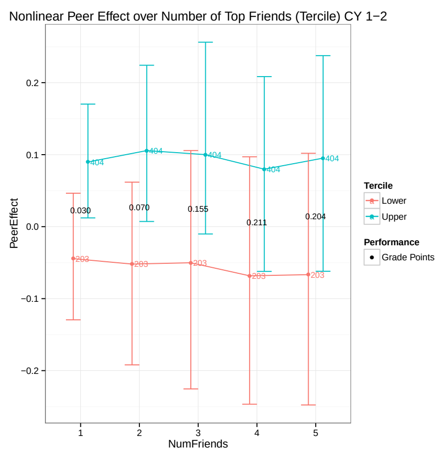
B.3.2 Terciles



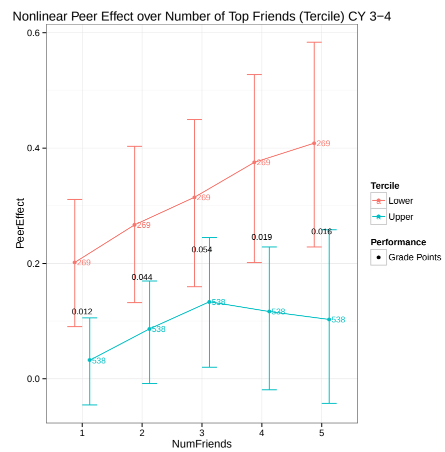
(a) No. Courses



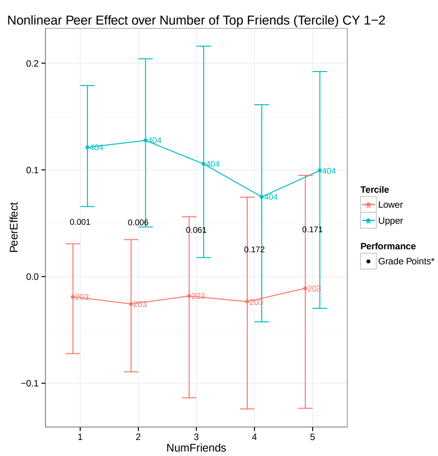
(b) No. Courses



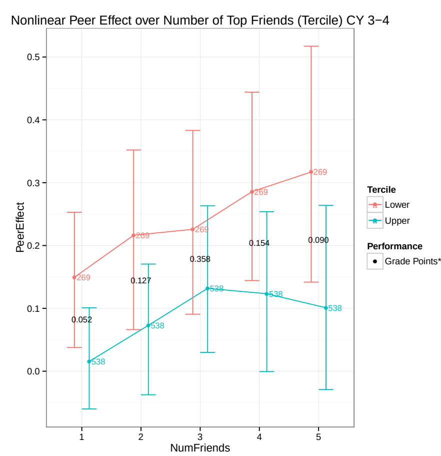
(c) Grade Points



(d) Grade Points



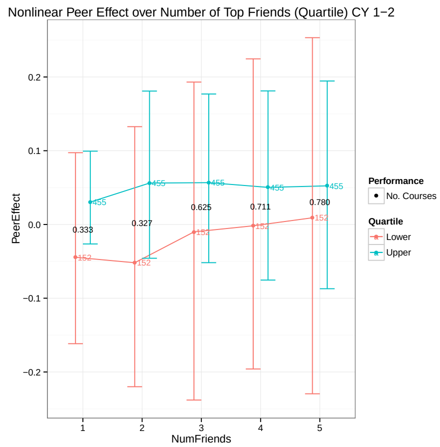
(e) Grade Points*



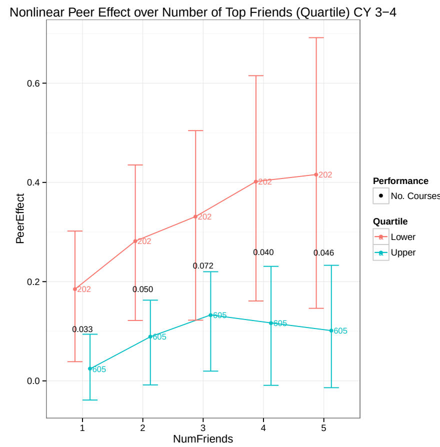
(f) Grade Points*

Figure B.8: Peer Effect by Terciles over No. Neighbors, split by Curricular Year.

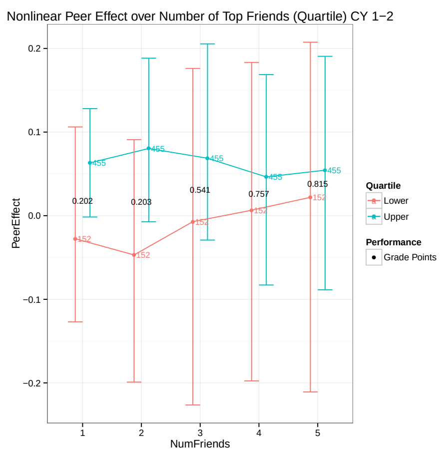
B.3.3 Quartiles



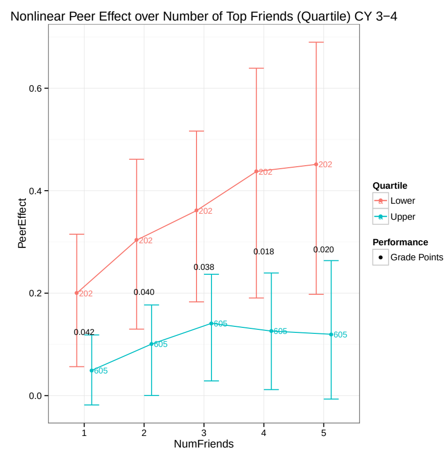
(a) No. Courses



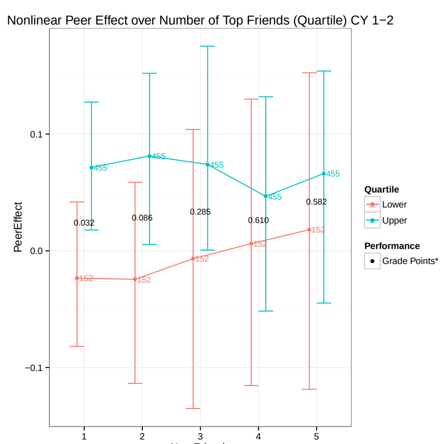
(b) No. Courses



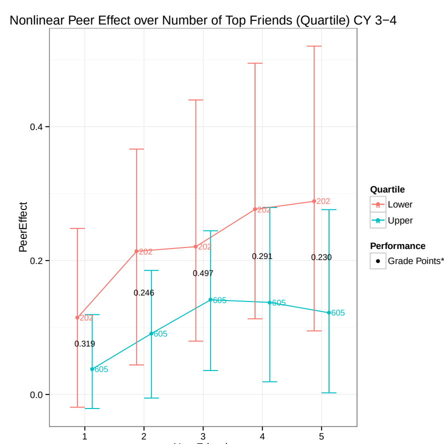
(c) Grade Points



(d) Grade Points



(e) Grade Points*

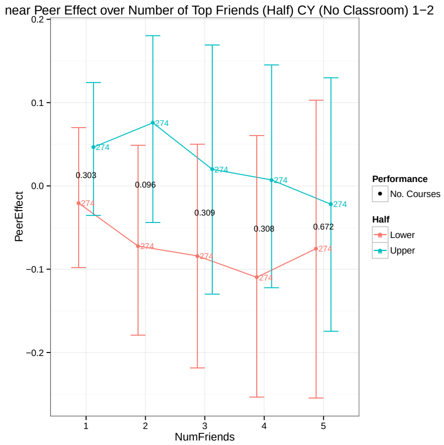


(f) Grade Points*

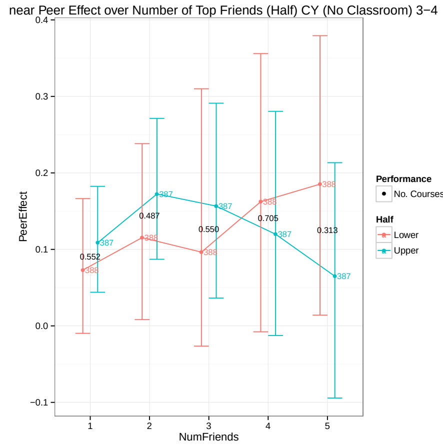
Figure B.9: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year.

B.4 Nonlinear Results without Classroom Usage (Spring 2007)

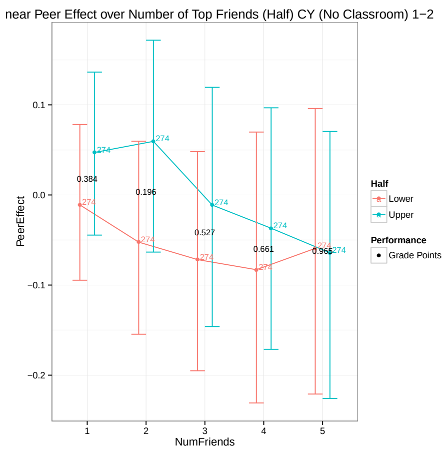
B.4.1 Halves



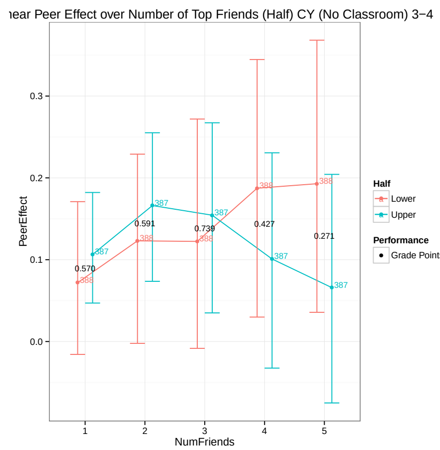
(a) No. Courses



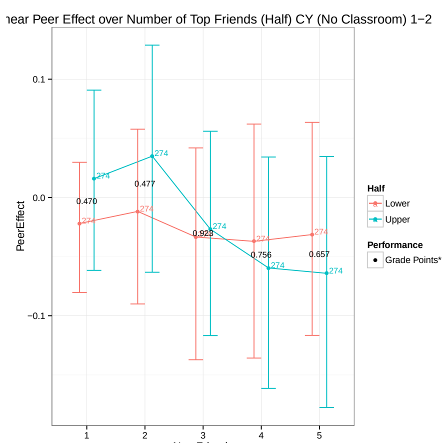
(b) No. Courses



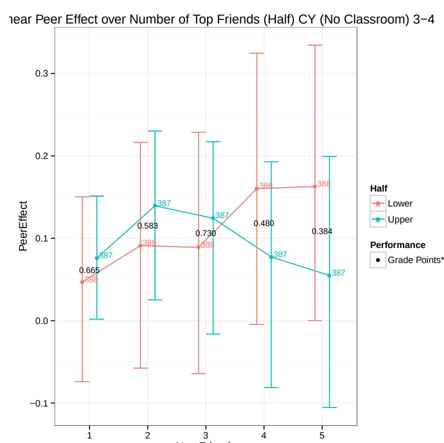
(c) Grade Points



(d) Grade Points



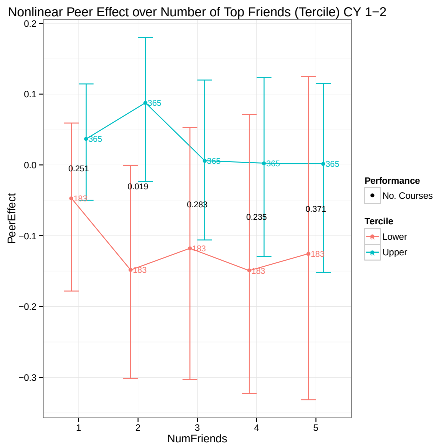
(e) Grade Points*



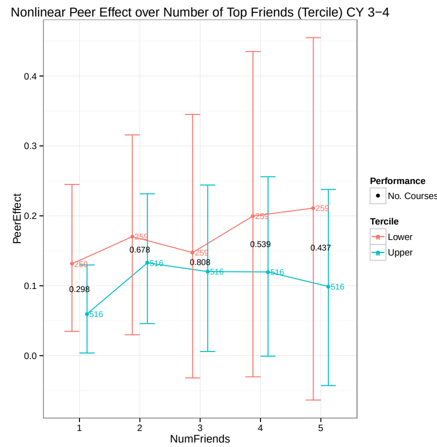
(f) Grade Points*

Figure B.10: Peer Effect by Halves over No. Neighbors, split by Curricular Year (No Classroom).

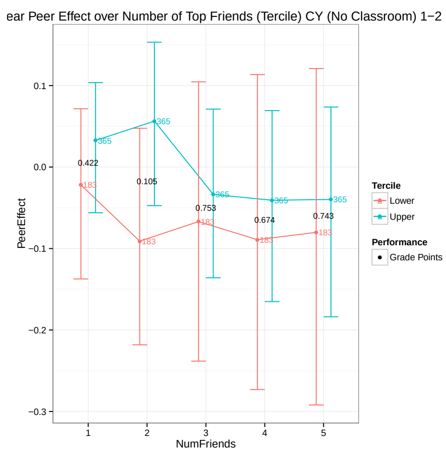
B.4.2 Terciles



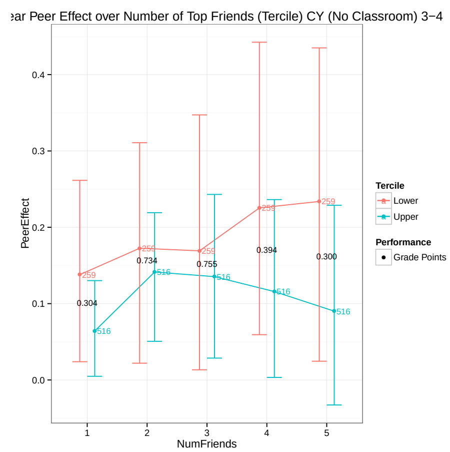
(a) No. Courses



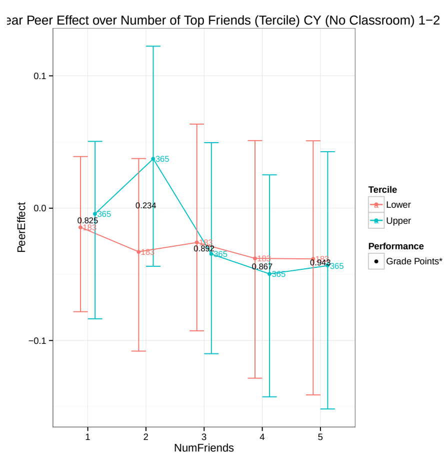
(b) No. Courses



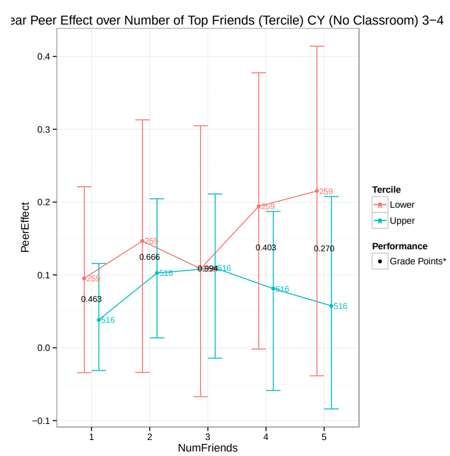
(c) Grade Points



(d) Grade Points



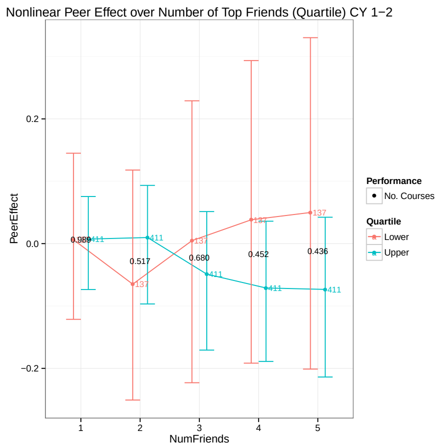
(e) Grade Points*



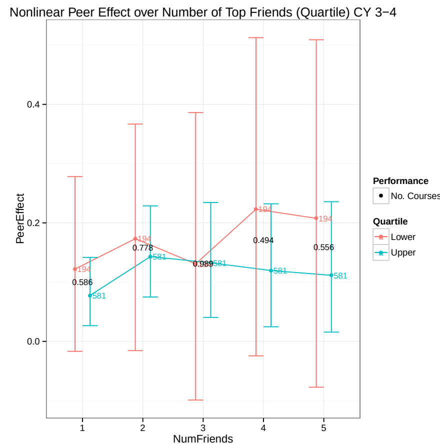
(f) Grade Points*

Figure B.11: Peer Effect by Terciles over No. Neighbors, split by Curricular Year (No Classroom).

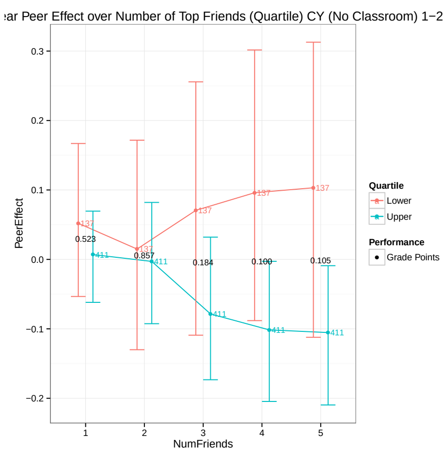
B.4.3 Quartiles



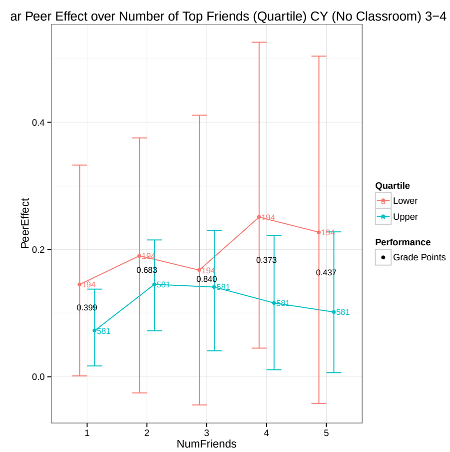
(a) No. Courses



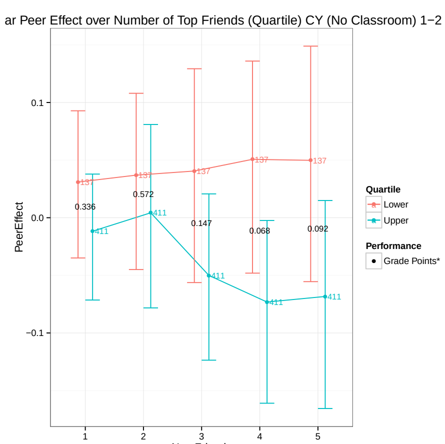
(b) No. Courses



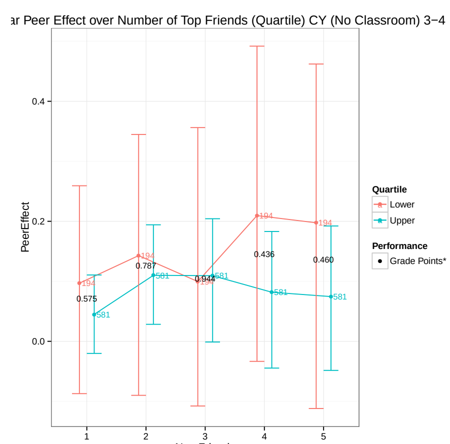
(c) Grade Points



(d) Grade Points



(e) Grade Points*

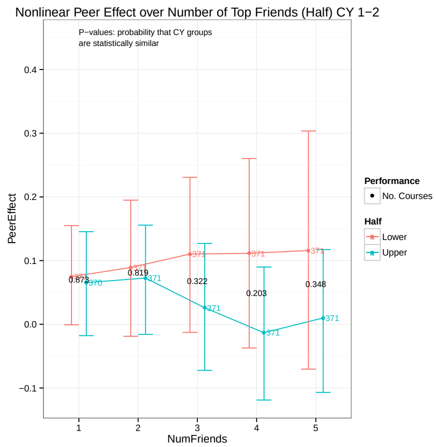


(f) Grade Points*

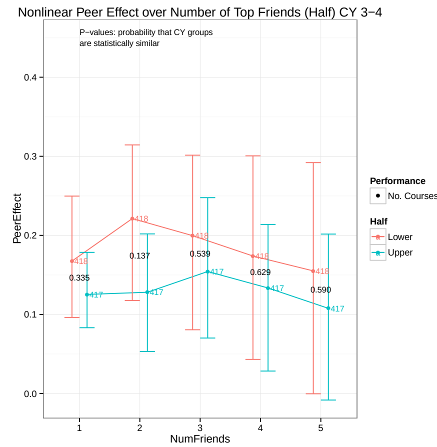
Figure B.12: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year (No Classroom).

B.5 Nonlinear Results with All Wifi Usage (Fall 2007)

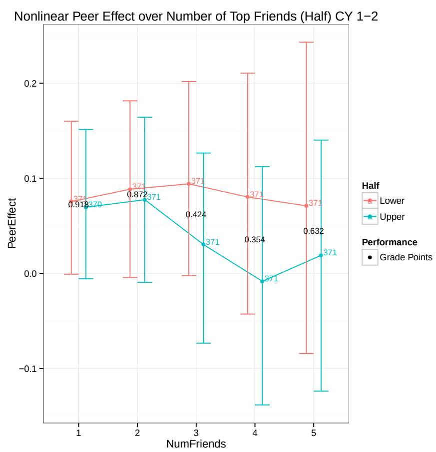
B.5.1 Halves



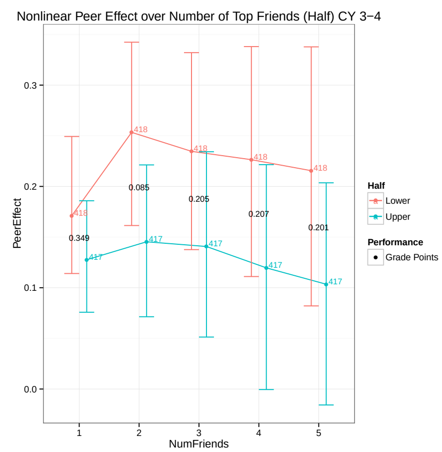
(a) No. Courses



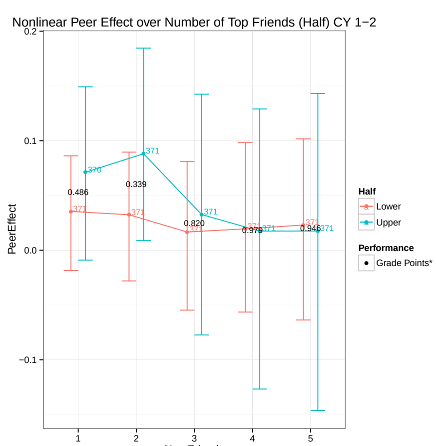
(b) No. Courses



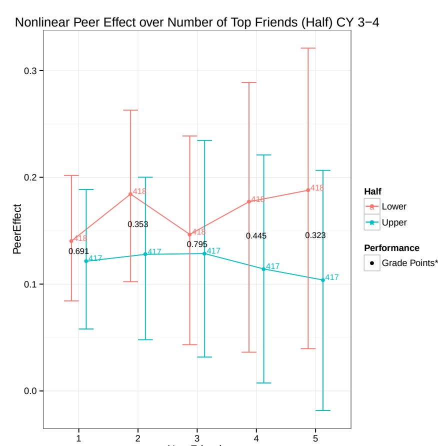
(c) Grade Points



(d) Grade Points



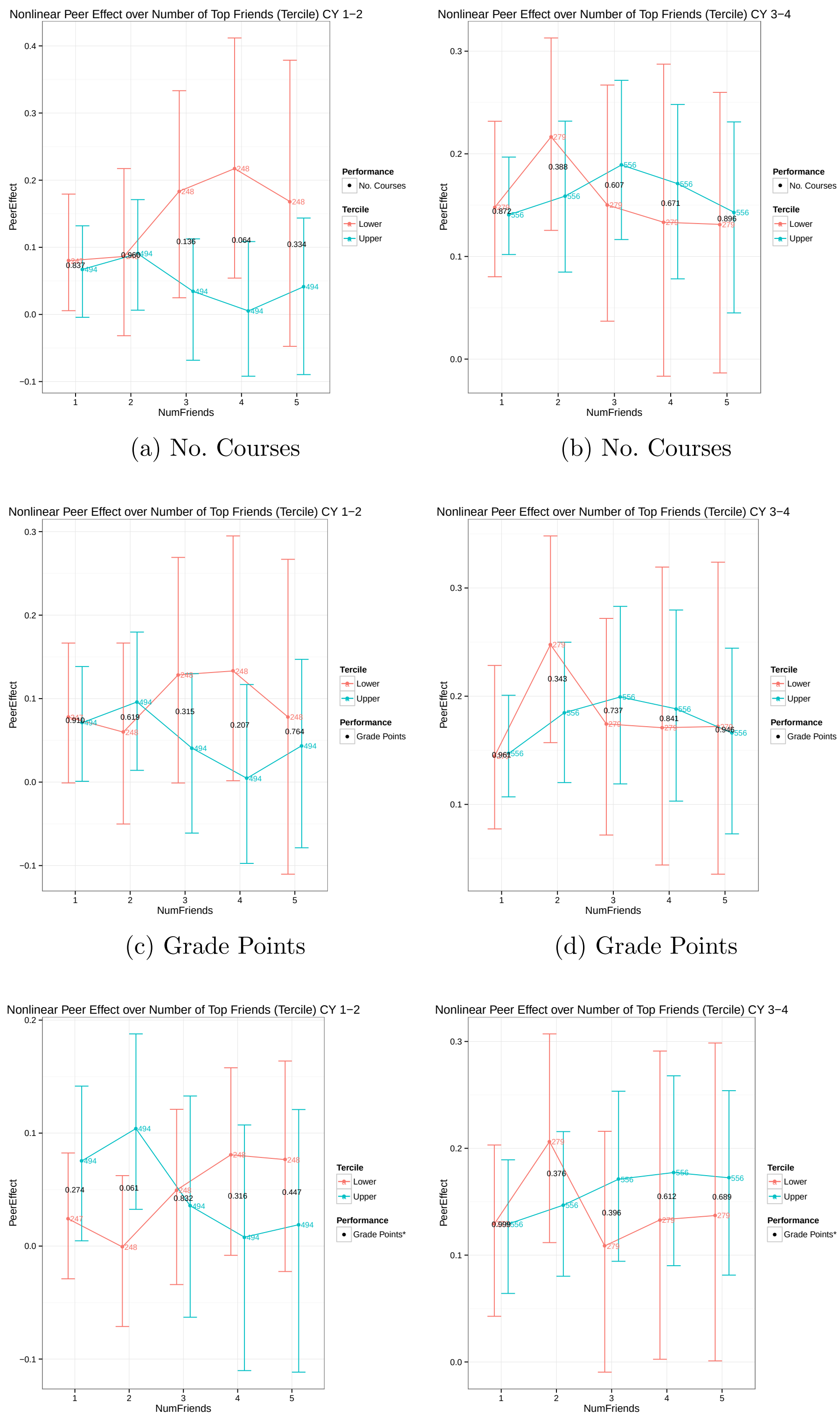
(e) Grade Points*



(f) Grade Points*

Figure B.13: Peer Effect by Halves over No. Neighbors, split by Curricular Year.

B.5.2 Terciles



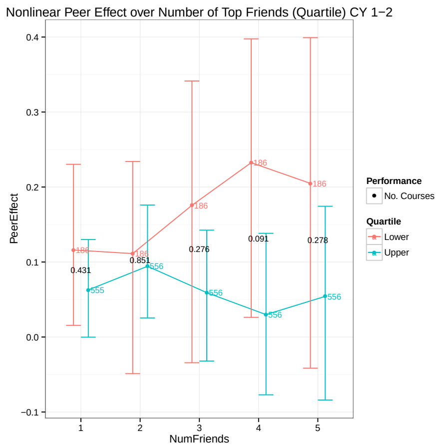
(a) No. Courses (b) No. Courses

(c) Grade Points (d) Grade Points

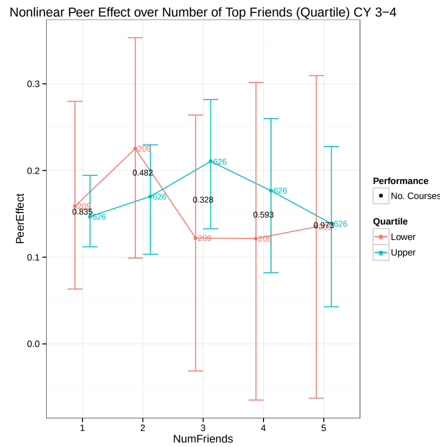
(e) Grade Points* (f) Grade Points*

Figure B.14: Peer Effect by Terciles over No. Neighbors, split by Curricular Year.

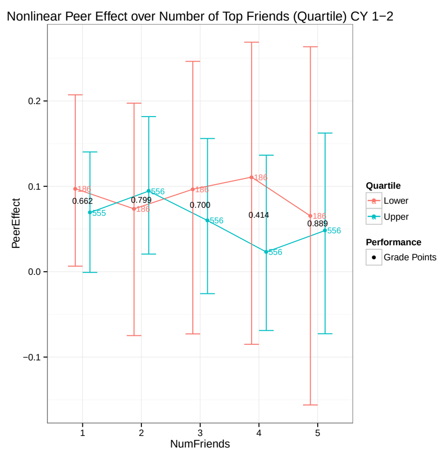
B.5.3 Quartiles



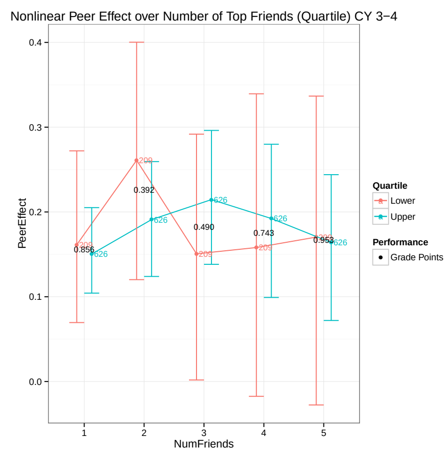
(a) No. Courses



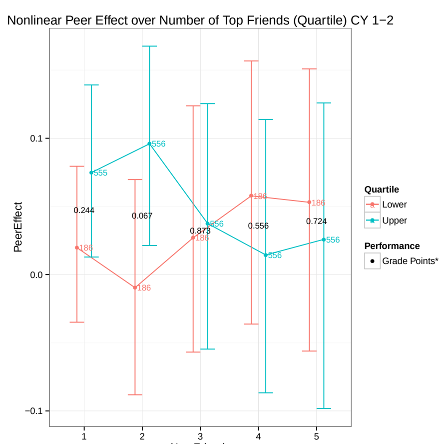
(b) No. Courses



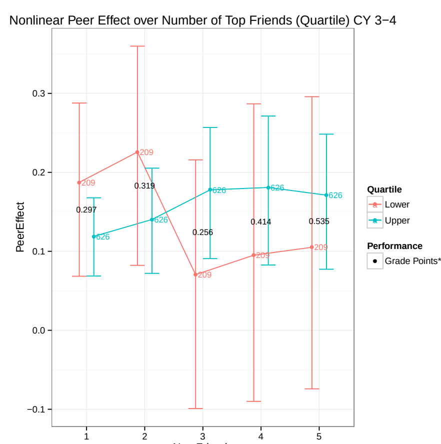
(c) Grade Points



(d) Grade Points



(e) Grade Points*

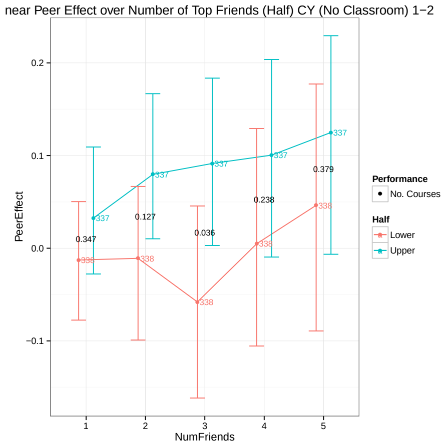


(f) Grade Points*

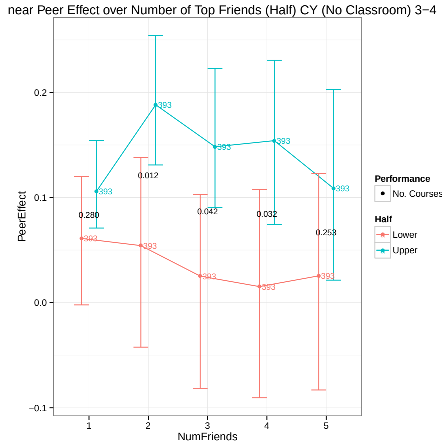
Figure B.15: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year.

B.6 Nonlinear Results without Classroom Usage (Fall 2007)

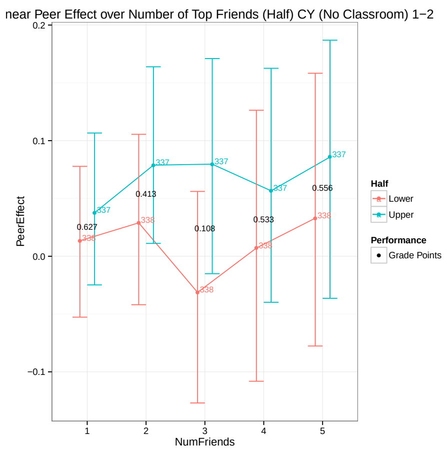
B.6.1 Halves



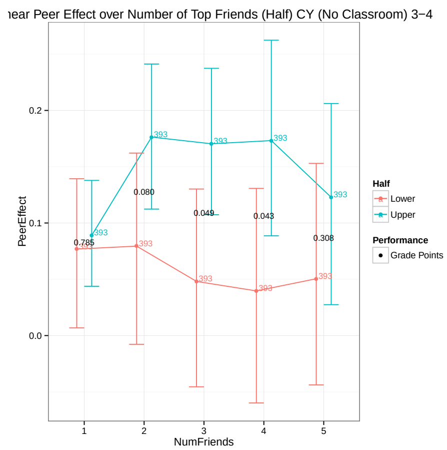
(a) No. Courses



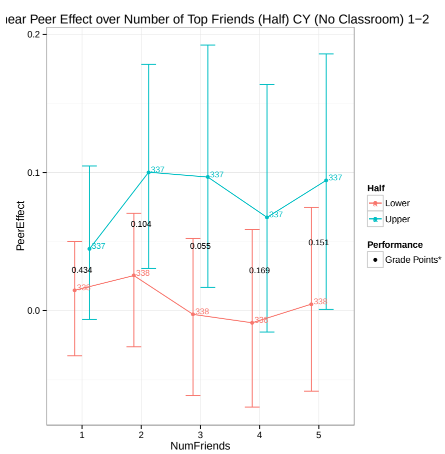
(b) No. Courses



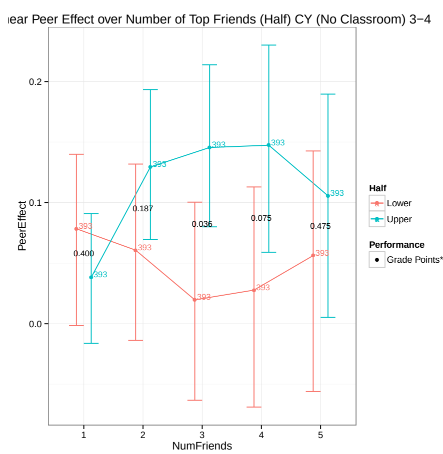
(c) Grade Points



(d) Grade Points



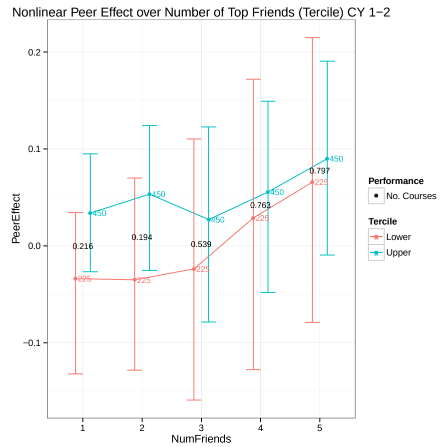
(e) Grade Points*



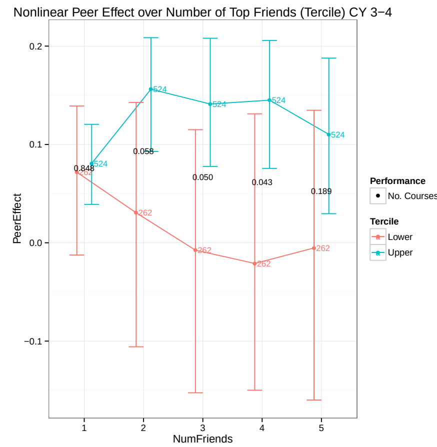
(f) Grade Points*

Figure B.16: Peer Effect by Halves over No. Neighbors, split by Curricular Year (No Classroom).

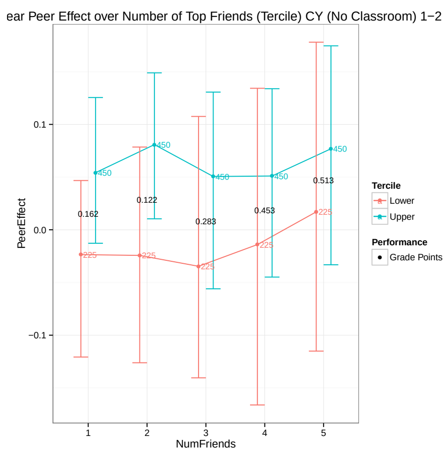
B.6.2 Terciles



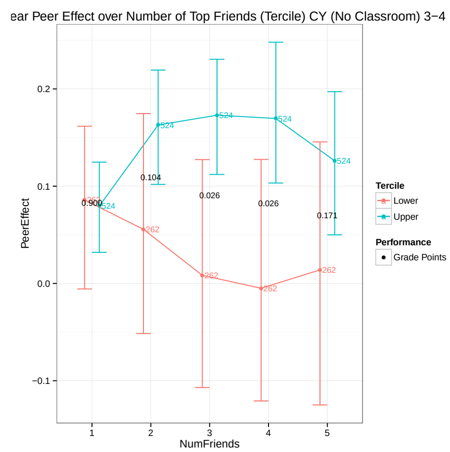
(a) No. Courses



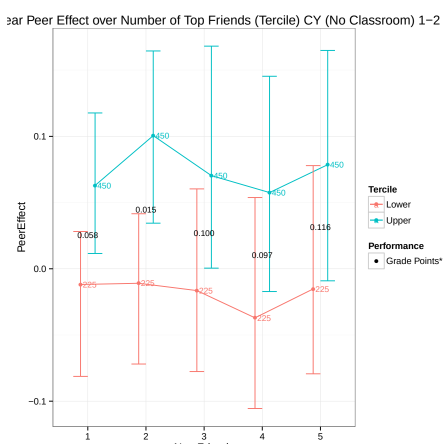
(b) No. Courses



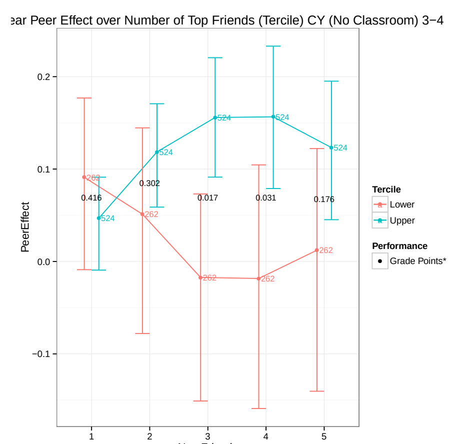
(c) Grade Points



(d) Grade Points



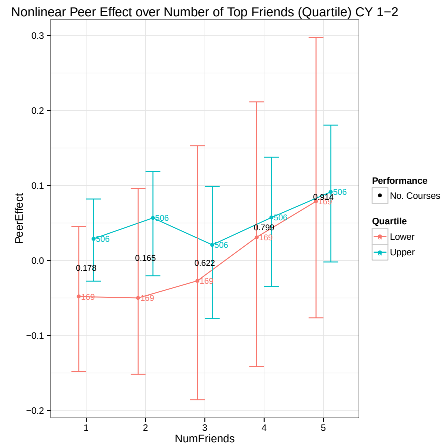
(e) Grade Points*



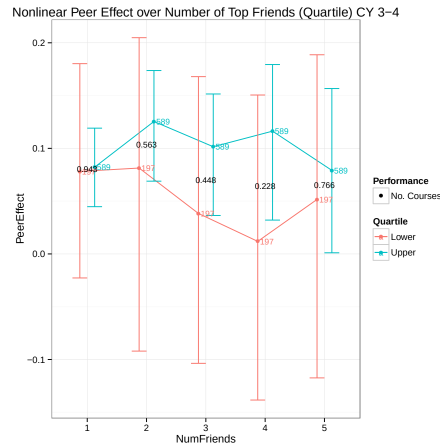
(f) Grade Points*

Figure B.17: Peer Effect by Terciles over No. Neighbors, split by Curricular Year (No Classroom).

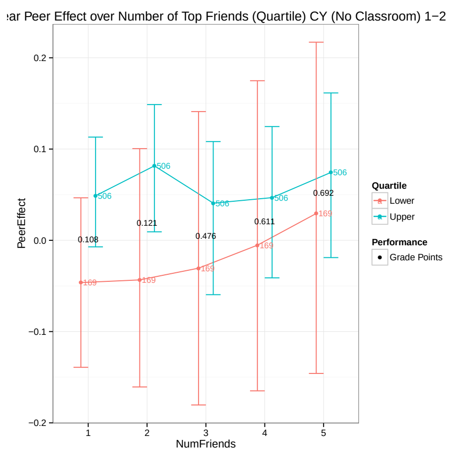
B.6.3 Quartiles



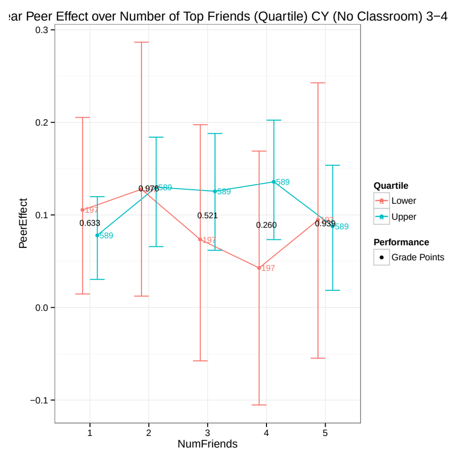
(a) No. Courses



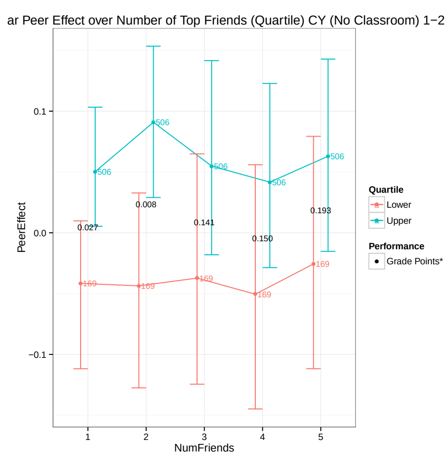
(b) No. Courses



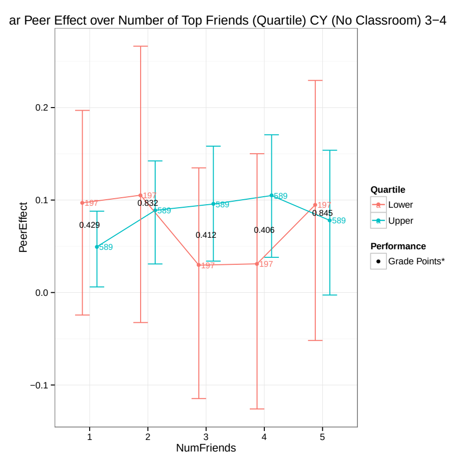
(c) Grade Points



(d) Grade Points



(e) Grade Points*

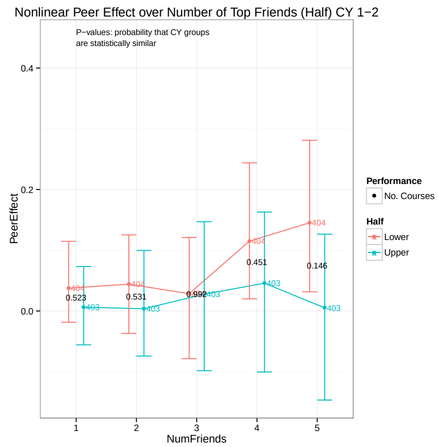


(f) Grade Points*

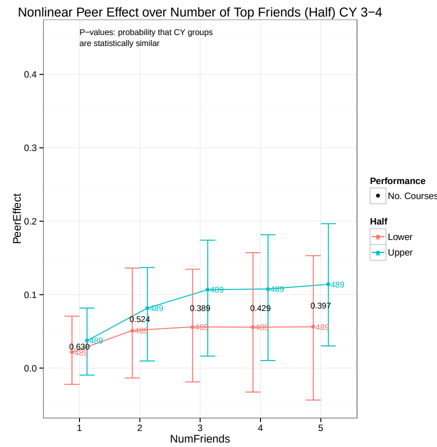
Figure B.18: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year (No Classroom).

B.7 Nonlinear Results with All Wifi Usage (Spring 2008)

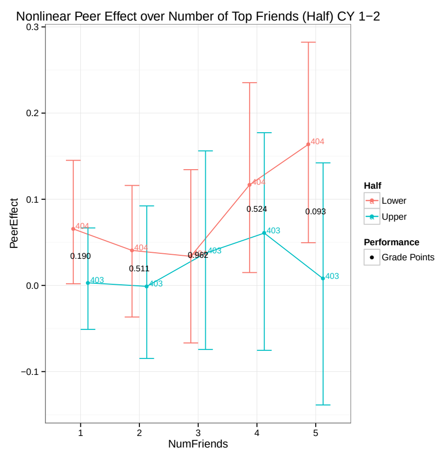
B.7.1 Halves



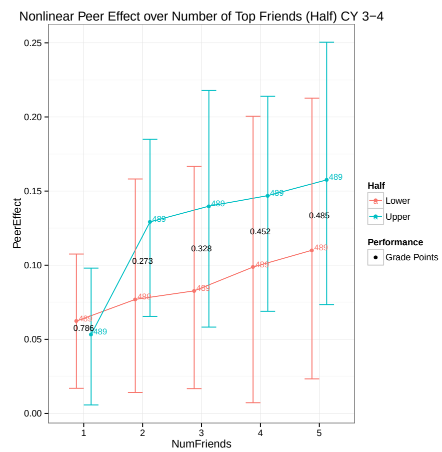
(a) No. Courses



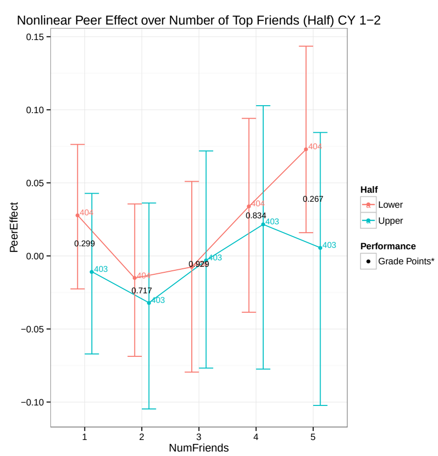
(b) No. Courses



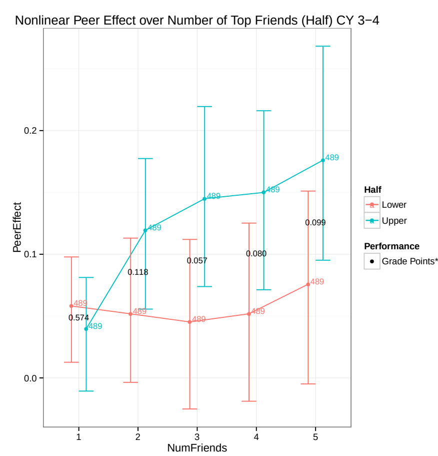
(c) Grade Points



(d) Grade Points



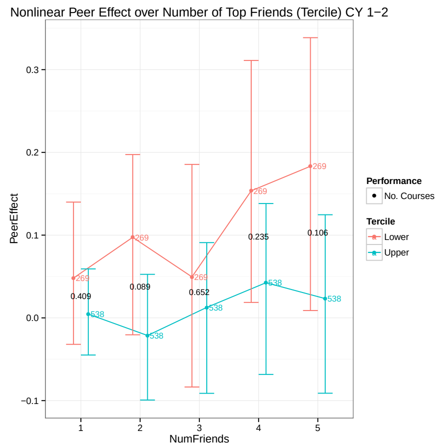
(e) Grade Points*



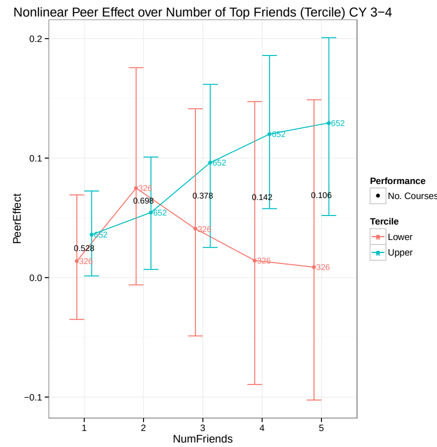
(f) Grade Points*

Figure B.19: Peer Effect by Halves over No. Neighbors, split by Curricular Year.

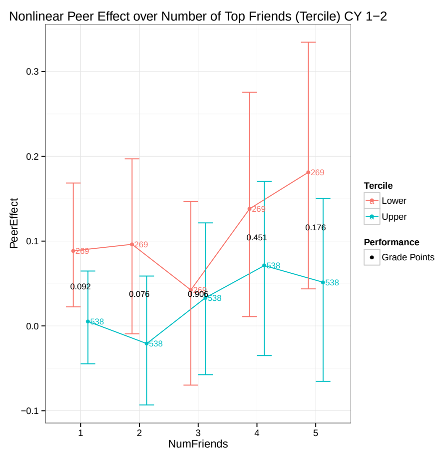
B.7.2 Terciles



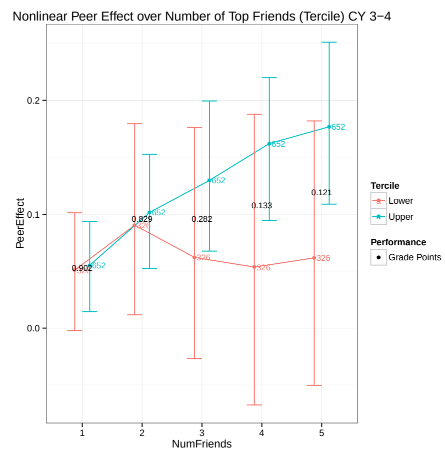
(a) No. Courses



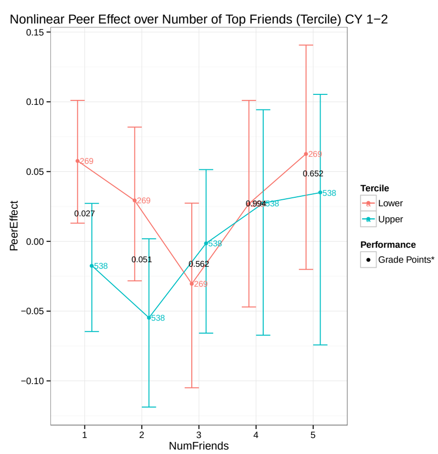
(b) No. Courses



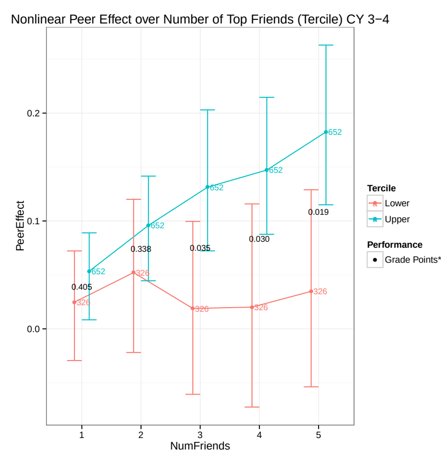
(c) Grade Points



(d) Grade Points



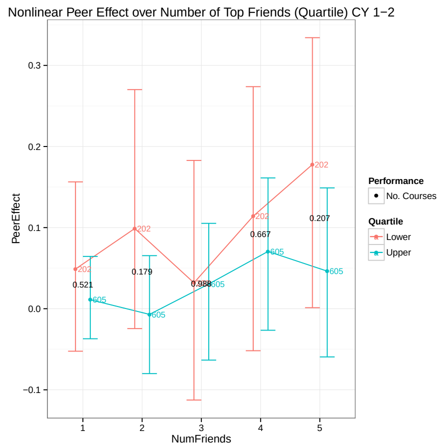
(e) Grade Points*



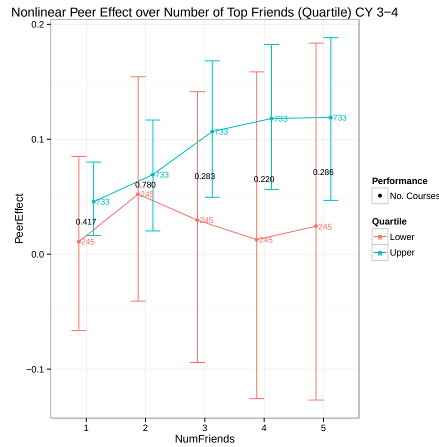
(f) Grade Points*

Figure B.20: Peer Effect by Terciles over No. Neighbors, split by Curricular Year.

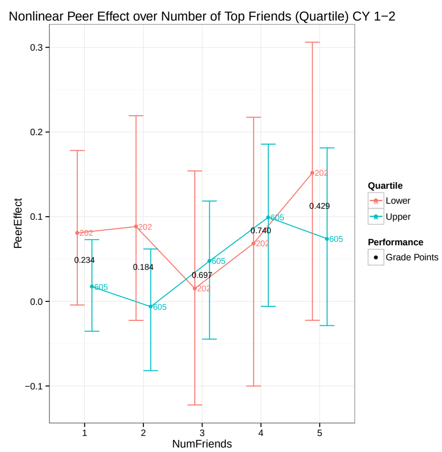
B.7.3 Quartiles



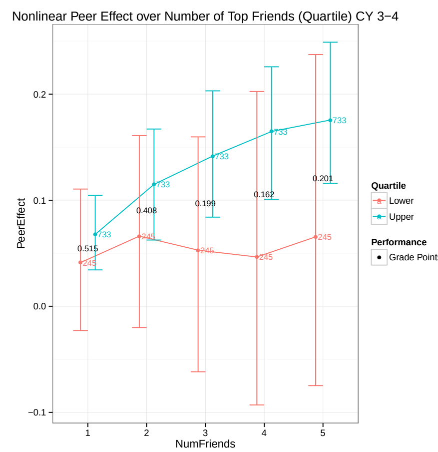
(a) No. Courses



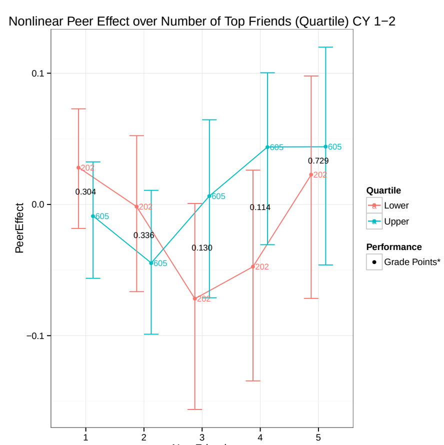
(b) No. Courses



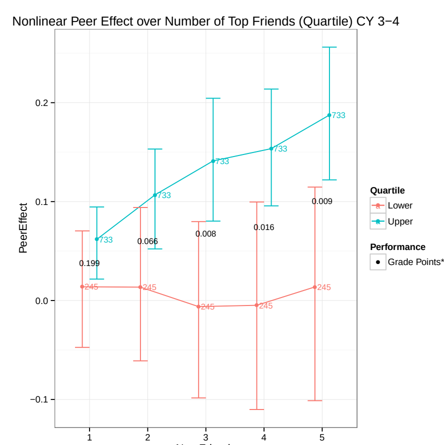
(c) Grade Points



(d) Grade Points



(e) Grade Points*

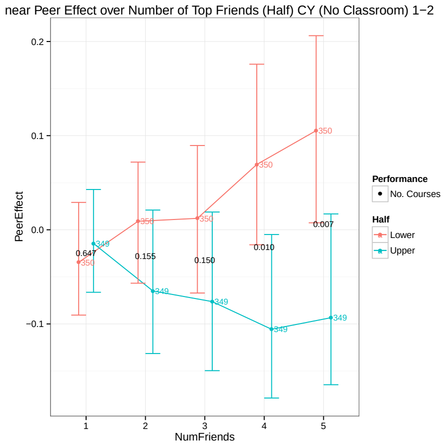


(f) Grade Points*

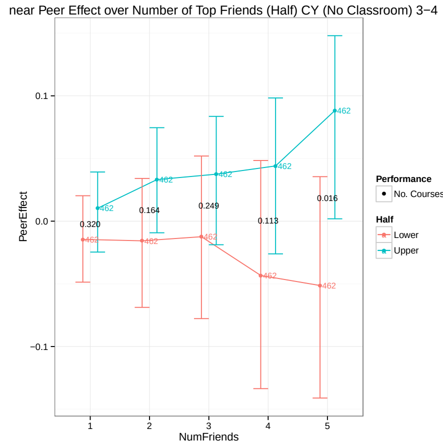
Figure B.21: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year.

B.8 Nonlinear Results without Classroom Usage (Spring 2008)

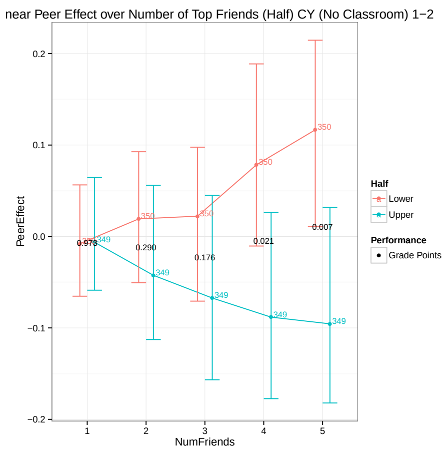
B.8.1 Halves



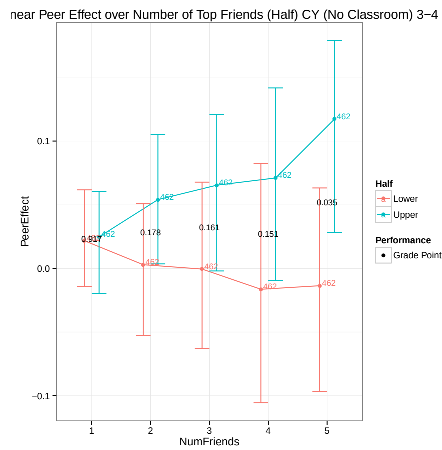
(a) No. Courses



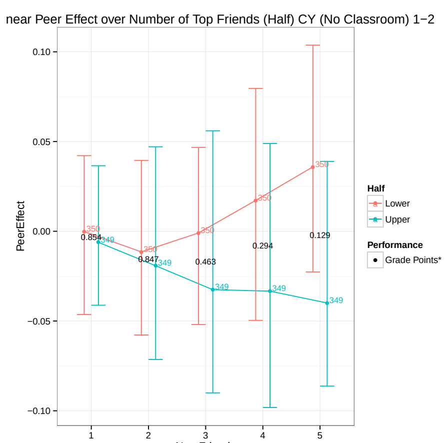
(b) No. Courses



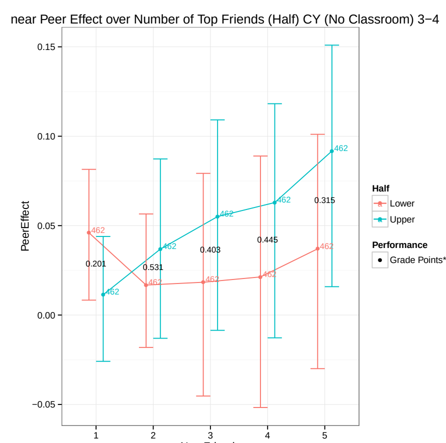
(c) Grade Points



(d) Grade Points



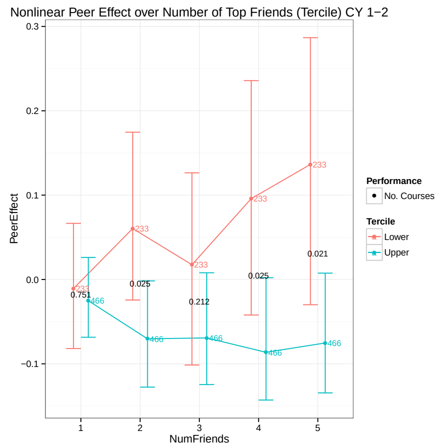
(e) Grade Points*



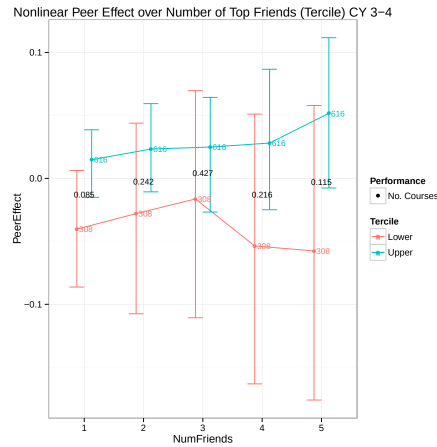
(f) Grade Points*

Figure B.22: Peer Effect by Halves over No. Neighbors, split by Curricular Year (No Classroom).

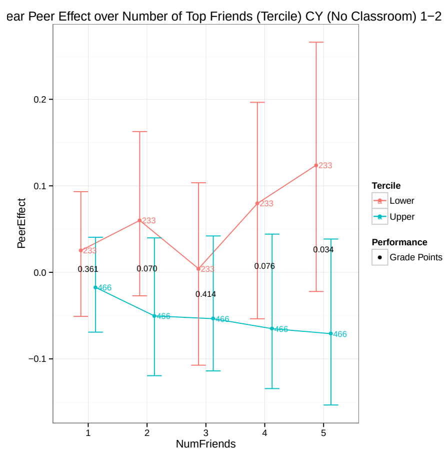
B.8.2 Terciles



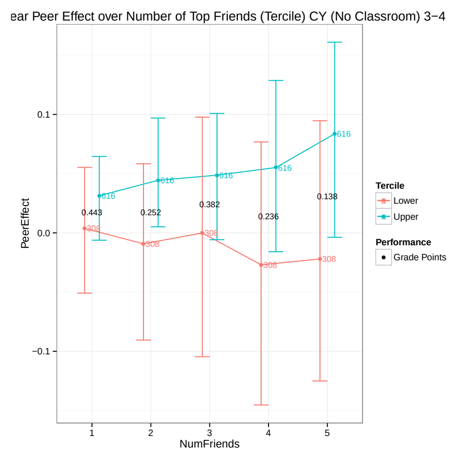
(a) No. Courses



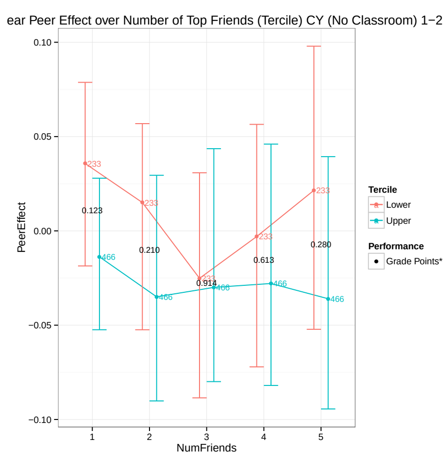
(b) No. Courses



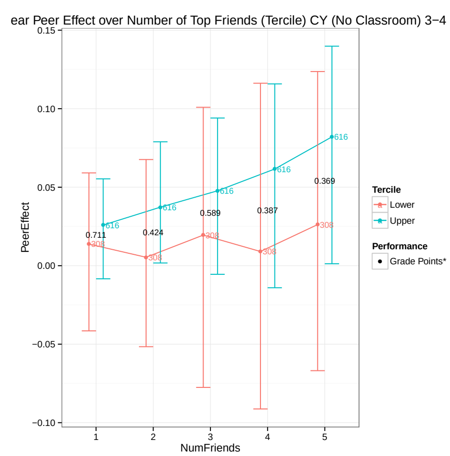
(c) Grade Points



(d) Grade Points



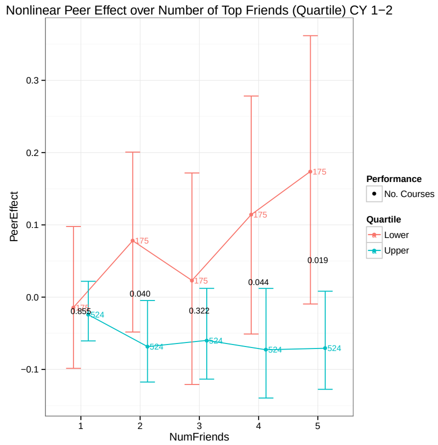
(e) Grade Points*



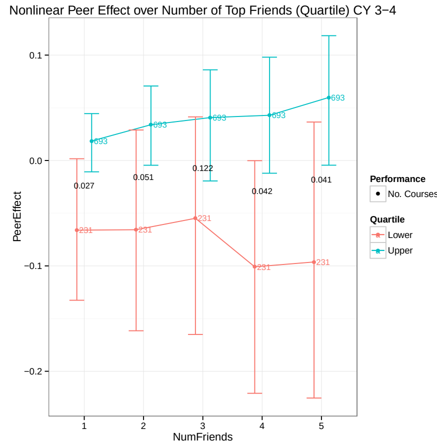
(f) Grade Points*

Figure B.23: Peer Effect by Terciles over No. Neighbors, split by Curricular Year (No Classroom).

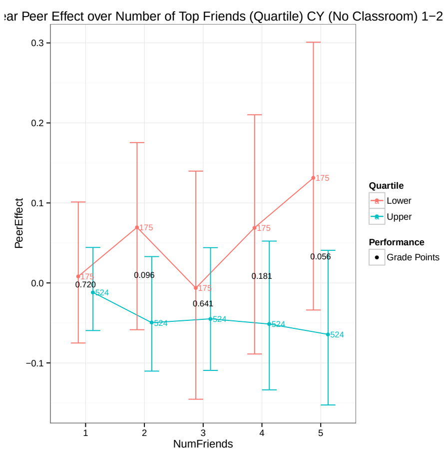
B.8.3 Quartiles



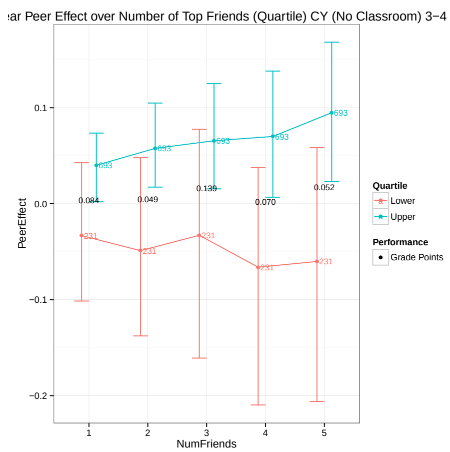
(a) No. Courses



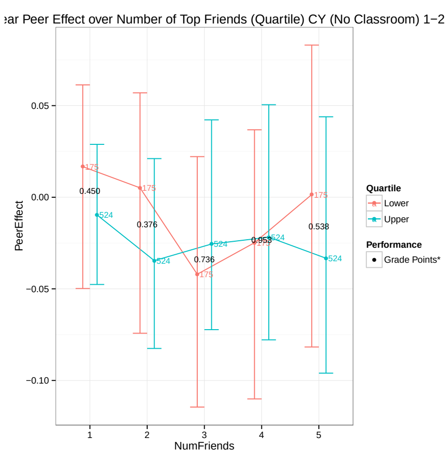
(b) No. Courses



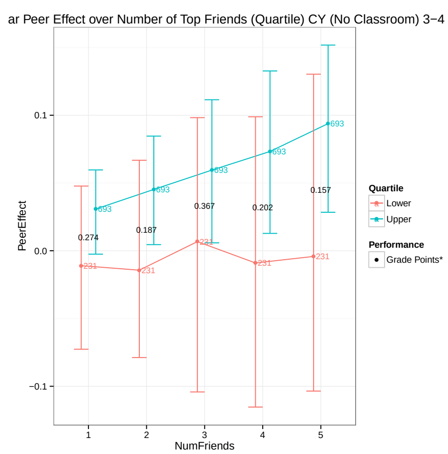
(c) Grade Points



(d) Grade Points



(e) Grade Points*

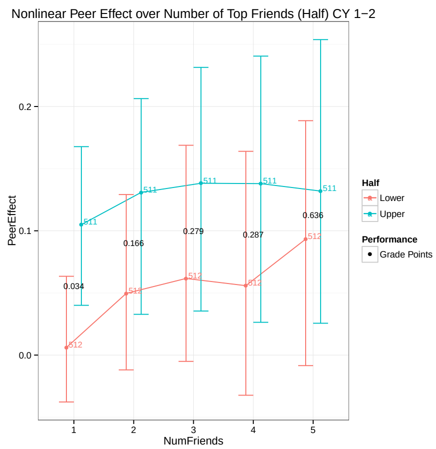


(f) Grade Points*

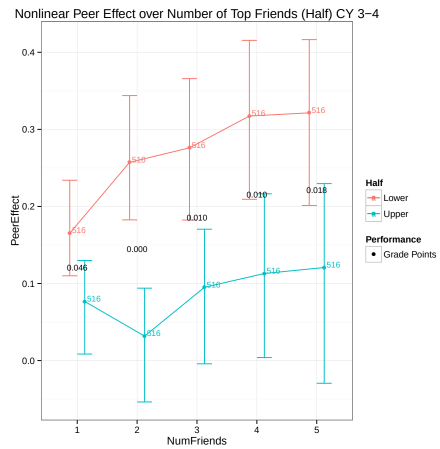
Figure B.24: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year (No Classroom).

B.9 Nonlinear Results with All Wifi Usage (Fall 2008)

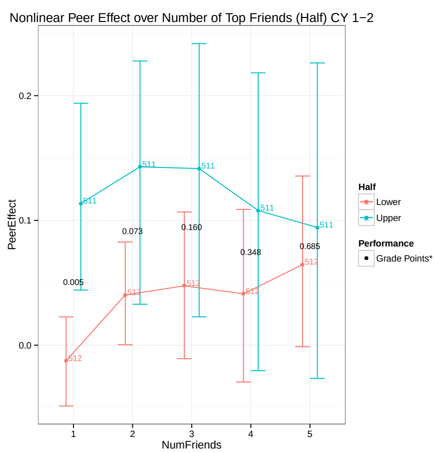
B.9.1 Halves



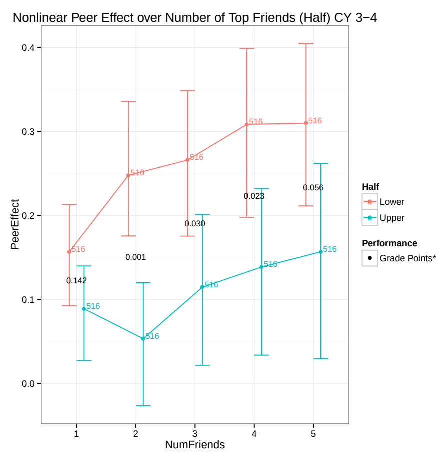
(a) Grade Points



(b) Grade Points



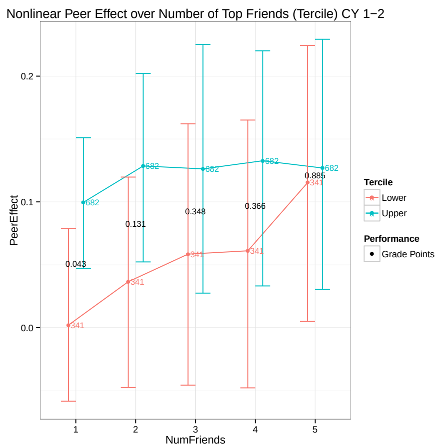
(c) Grade Points*



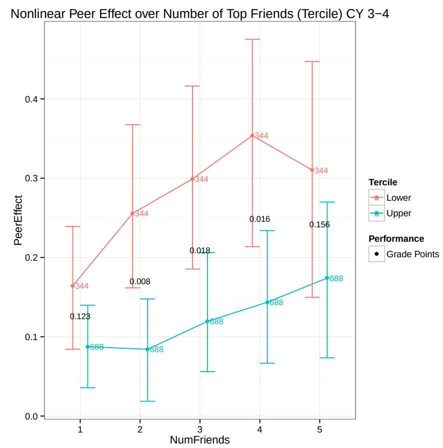
(d) Grade Points*

Figure B.25: Peer Effect by Halves over No. Neighbors, split by Curricular Year.

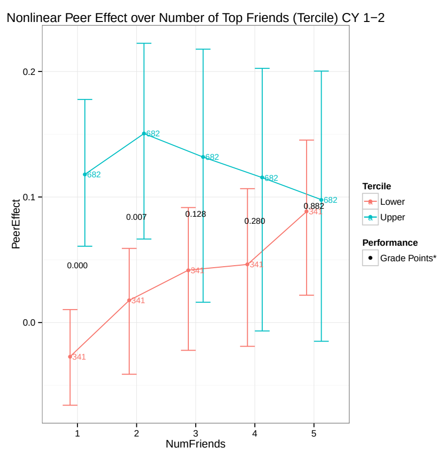
B.9.2 Terciles



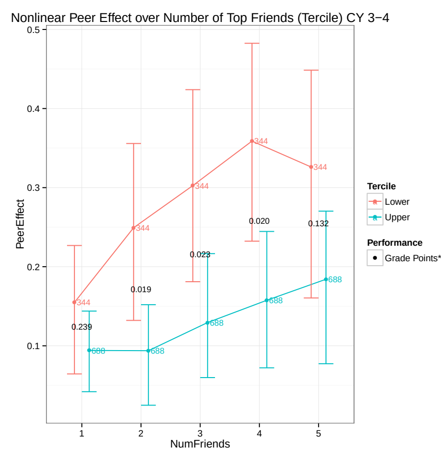
(a) Grade Points



(b) Grade Points



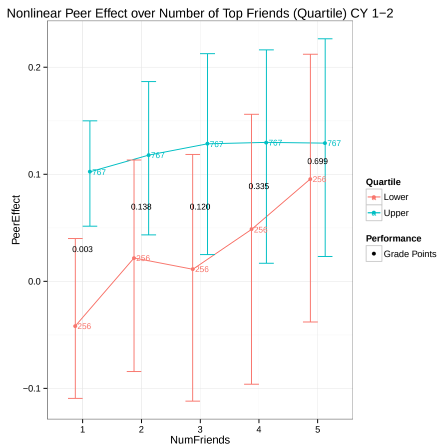
(c) Grade Points*



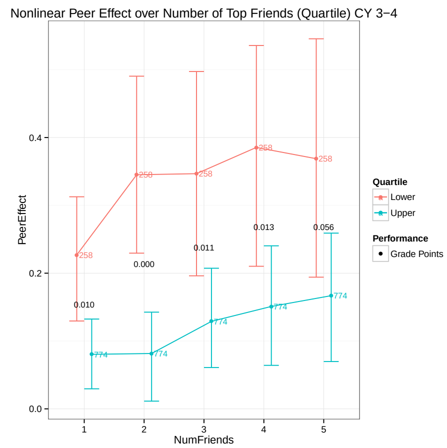
(d) Grade Points*

Figure B.26: Peer Effect by Terciles over No. Neighbors, split by Curricular Year.

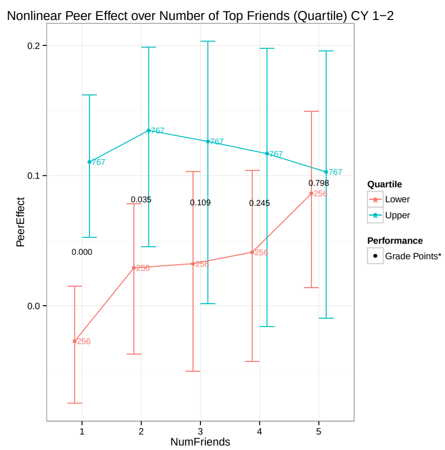
B.9.3 Quartiles



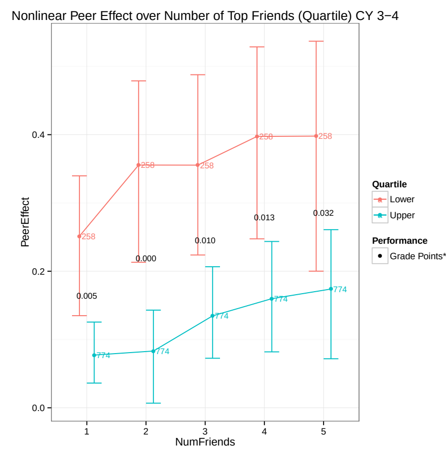
(a) Grade Points



(b) Grade Points



(c) Grade Points*

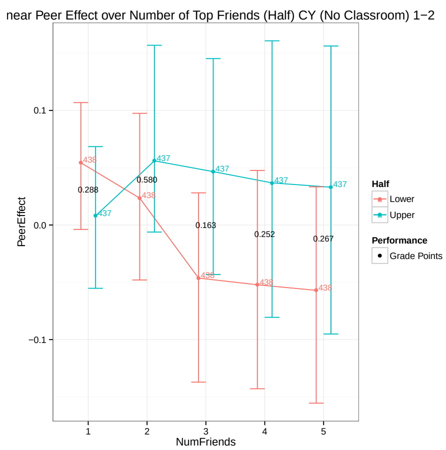


(d) Grade Points*

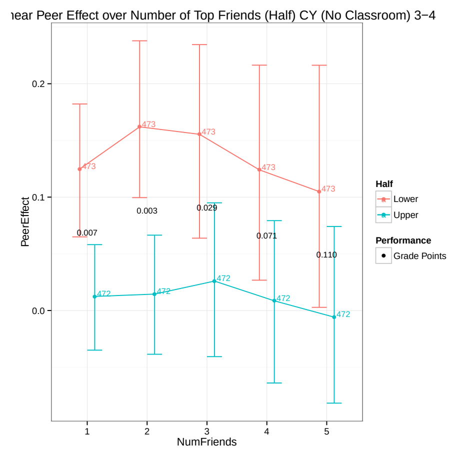
Figure B.27: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year.

B.10 Nonlinear Results without Classroom Usage (Fall 2008)

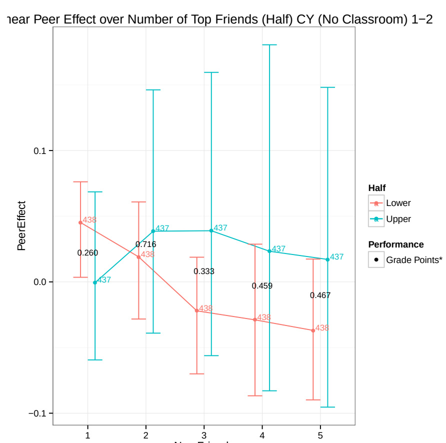
B.10.1 Halves



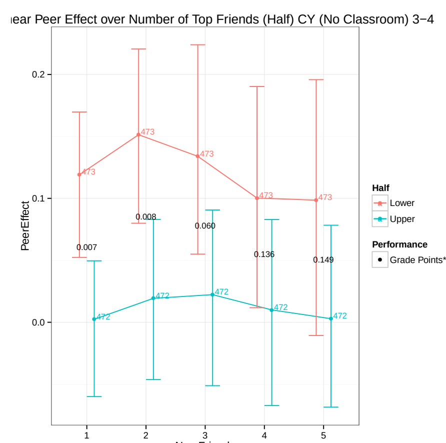
(a) Grade Points



(b) Grade Points



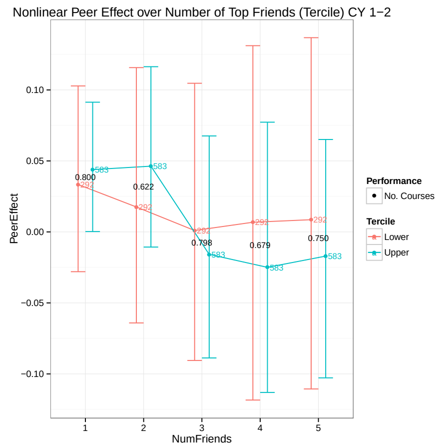
(c) Grade Points*



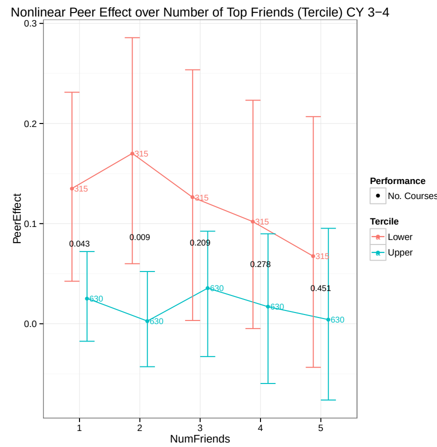
(d) Grade Points*

Figure B.28: Peer Effect by Halves over No. Neighbors, split by Curricular Year (No Classroom).

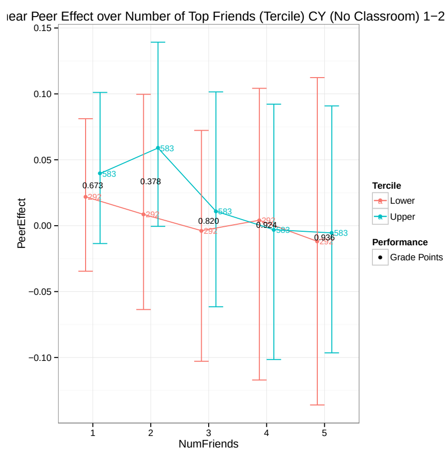
B.10.2 Terciles



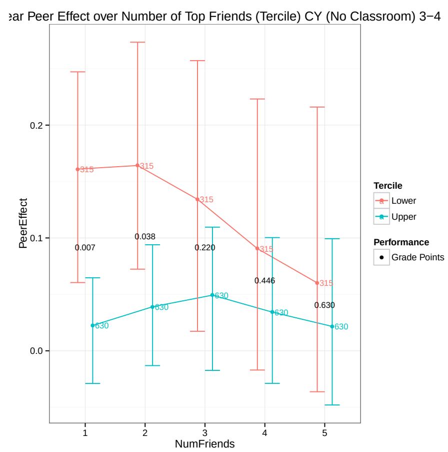
(a) No. Courses



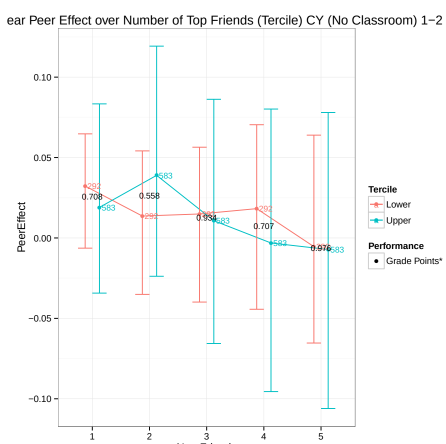
(b) No. Courses



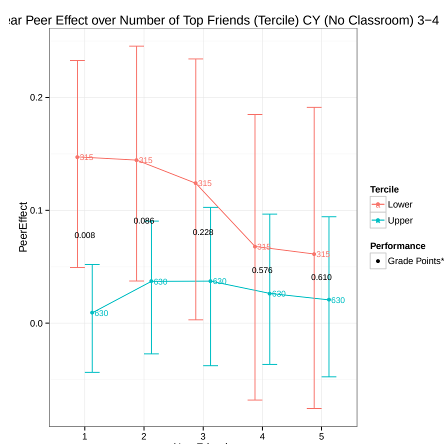
(c) Grade Points



(d) Grade Points



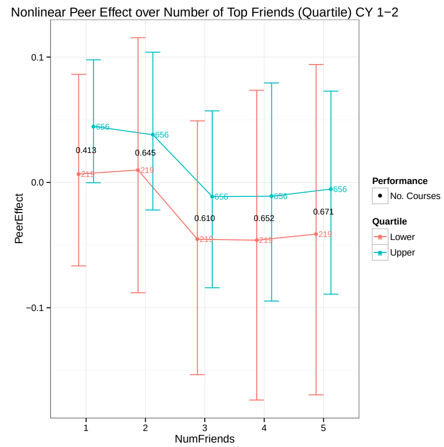
(e) Grade Points*



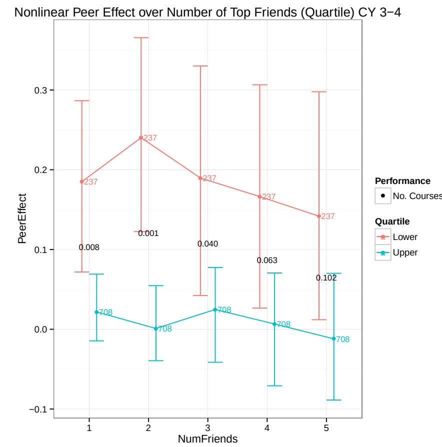
(f) Grade Points*

Figure B.29: Peer Effect by Terciles over No. Neighbors, split by Curricular Year (No Classroom).

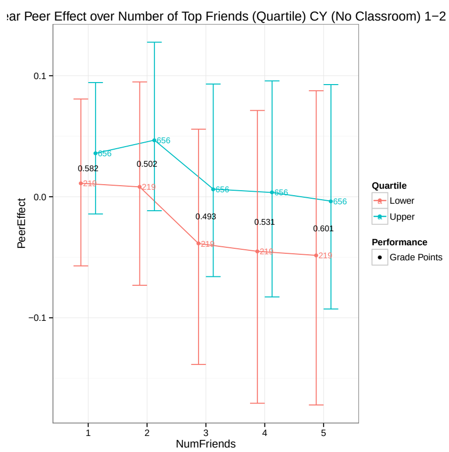
B.10.3 Quartiles



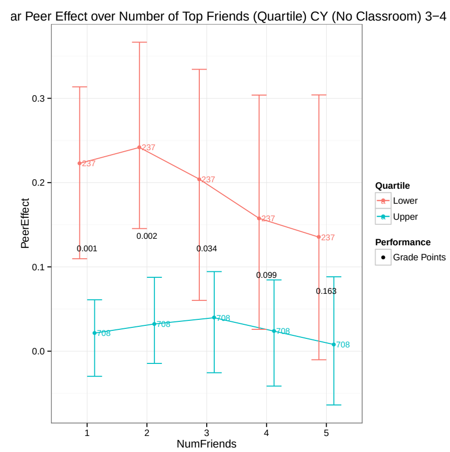
(a) No. Courses



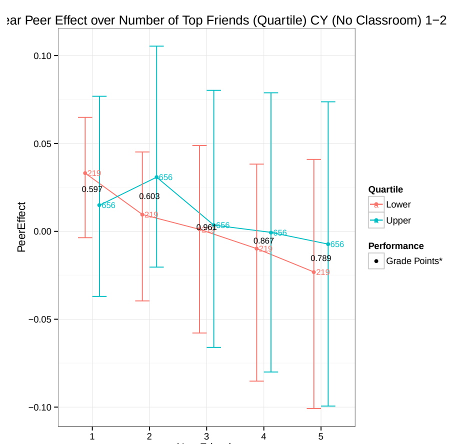
(b) No. Courses



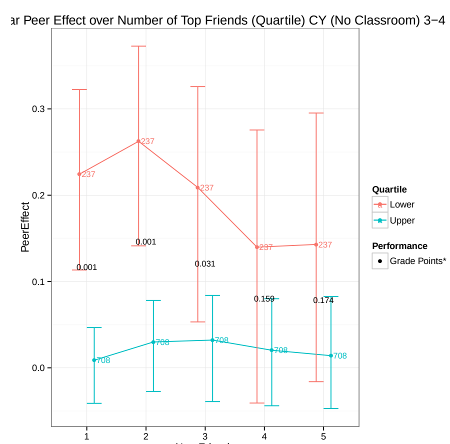
(c) Grade Points



(d) Grade Points



(e) Grade Points*



(f) Grade Points*

Figure B.30: Peer Effect by Quartiles over No. Neighbors, split by Curricular Year (No Classroom).

Appendix C

Common Neighbors

Figure C.1 shows the average percentage of overlap in the Top N Neighbors across social networks with all wifi data and without classroom data. Although the percentage appears very low, note that wifi usage follows a power law distribution (with few users constituting the majority of the usage), so what is important here is that we see non-zero overlap.

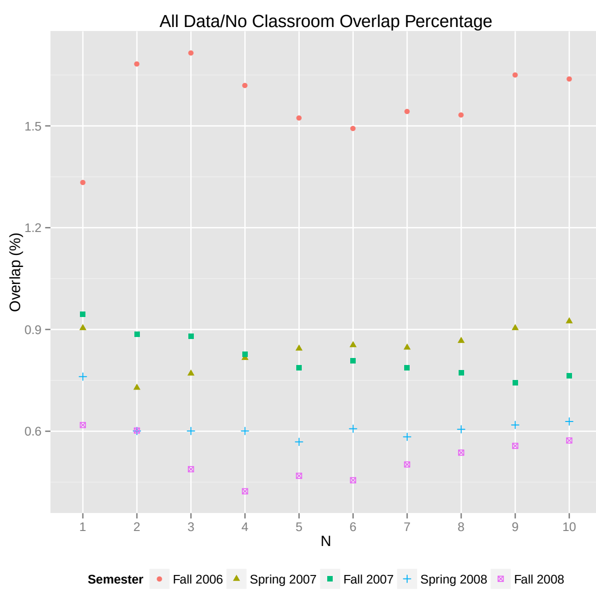


Figure C.1: Common Neighbors with and without Classroom Data

References

Allen, J. D. (2005). Grades as Valid Measures of Academic Achievement of Classroom Learning. *The Clearing House*, *78*(5), 218–223. doi: doi:10.3200/TCHS.78.5.218-223

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008, August). Influence and Correlation in Social Networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 7–15). doi: doi:10.1145/1401890.1401897

Angrist, J., & Lavy, V. (2002). New Evidence on Classroom Computers and Pupil Learning. *The Economic Journal*, *112*(482), 735–765. doi: doi:10.1111/1468-0297.00068/full

Arabasz, P., & Pirani, J. (2002a). *Wireless Networking at Dartmouth College [Case Study 9]*. Boulder, CO: EDUCAUSE Center for Applied Research. Retrieved from https://net.educause.edu/ir/library/pdf/ers0202/cs/ecs0209.pdf

Arabasz, P., & Pirani, J. (2002b). *Wireless Networking in Higher Education* (Vol. 2). Boulder, CO: EDUCAUSE Center for Applied Research.

Aral, S., Muchnik, L., & Sundararajan, A. (2009, December). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*(51), 21544–21549. doi: doi:10.1073/pnas.0908800106

Awwad, F., Ayesh, A., & Awwad, S. (2013). Are Laptops Distracting Educational Tools in Classrooms. *Procedia Social and Behavioral Sciences*, *103*, 154–160. doi: doi:10.1016/j.sbspro.2013.10.320

Belo, R., & Ferreira, P. (2012, September). Using Randomization to Identify Social Influence in Mobile Networks. In *2012 International Conference on Privacy, Security, Risk and Trust (PASSAT)* (pp. 599–604). doi: doi:10.1109/SocialCom-PASSAT.2012.62

Belo, R., & Ferreira, P. (2013, June). Is Social Influence Always Positive? Evidence from a Very Large Mobile Network. In *Statistical challenges in ecommerce research symposium.* Lisbon.

Belo, R., Ferreira, P., & Telang, R. (2014, February). Broadband in School: Impact on Student Performance. *Management Science*, *60*(2), 265–282. doi: doi:10.1287/mnsc.2013.1770

Ben Youssef, A., & Dahmani, M. (2008). The Impact of ICT on Student Performance in Higher Education: Direct Effects, Indirect Effects and Organizational Change. In *The economics of e-learning.* Revista de Universidad y Sociedad del Conocimiento (RUSC). Retrieved from http://www.uoc.edu/ojs/index.php/rusc/article/view/v5n1-ben-youssef-dahmani

Burke, M. A., & Sass, T. R. (2013). Classroom Peer Effects and Student Achievement. *Journal of Labor Economics*, *31*(1), 51–82. doi: doi:10.1086/666653

Calvó-Armengol, A., Patacchini, E., & Zenou, Y. (2009, October). Peer Effects and Social Networks in Education. *Review of Economic Studies*, *76*(4), 1239–1267. doi: doi:10.1111/j.1467-937X.2009.00550.x

Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, *27*(3), 439–464.

Chen, E. (2006). *Laptops nixed in some law classes.* Retrieved from http://www.thedp.com/article/2006/04/laptops_nixed_in_some_law_classes

Council of Economic Advisors. (2011, September). *Unleashing the Potential of Educational Technology.* Washington, D.C.: Executive Office of the President.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research.

InterJournal, *Complex Systems*, 1695. Retrieved from http://igraph.sf.net

DiNicola, M. (2004). *Pathological Internet Use Among College Students: The Prevalance of Pathological Internet Use and Its Correlates.* Unpublished doctoral dissertation, Ohio University.

Englander, F., Terregrossa, R. A., & Wang, Z. (2010, February). Internet use among college students: tool or toy? *Educational Review*, *62*(1), 85–96. doi: doi:10.1080/00131910903519793

Epple, D., & Romano, R. E. (2011). Peer Effects in Education: A Survey of the Theory and Evidence. In *Handbook of Social Economics* (pp. 1053–1163). Elsevier B.V. doi: doi:10.1016/B978-0-444-53707-2.00003-7

European Higher Education Area. (1999). *The Bologna Declaration of 19 June 1999* (Tech. Rep.).

Foster, G. (2006, September). It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics*, *90*(8-9), 1455–1475. doi: doi:10.1016/j.jpubeco.2005.12.001

Fried, C. (2008, April). In-class laptop use and its effects on student learning. *Computers & Education*, *50*(3), 906–914. doi: doi:10.1016/j.compedu.2006.09.006

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011, May). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, *42*(8), 1–28. Retrieved from http://www.jstatsoft.org/v42/i08/

Imai, K., King, G., & Lau, O. (2007). *Zelig: Everyone's Statistical Software.* Retrieved from http://gking.harvard.edu/zelig

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistics Society A*, *171*(2), 481–502. Retrieved from http://imai.princeton.edu/research/files/matchse.pdf

Jain, T., & Langer, N. (2013, October). *Does who you know matter? Unraveling the influence of student networks on academic performance.* Working Paper. Retrieved from http://www.sole-jole.org/14097.pdf

Kubey, R. W., Lavin, M. J., & Barrows, J. R. (2001). Internet Use and Collegiate Academic Performance Decrements: Early findings. *Journal of Communication*, *51*(2), 366–382.

La Fond, T., & Neville, J. (2010, April). Randomization Tests for Distinguishing Social Influence and Homophily Effects. In *Proceedings of the 19th international conference on world wide web.* ACM. doi: doi:10.1145/1772690.1772752

Machin, S., McNally, S., & Silva, O. (2007, July). New Technology in Schools: Is There a Payoff? *The Economic Journal*, *117*(522), 1145–1167. Retrieved from http://doi.wiley.com/10.1111/j.1468-0297.2007.02070.x doi: doi:10.1111/j.1468-0297.2007.02070.x

Malamud, O., & Pop-Eleches, C. (2011, July). Home Computer Use and the Development of Human Capital. *Quarterly Journal of Economics*, *126*(2), 987–1027. doi: doi:10.1093/qje/qjr008

Marmaros, D., & Sacerdote, B. (2006, February). How Do Friendships Form? *Quarterly Journal of Economics*, *121*(1), 79–119. doi: doi:10.1093/qje/121.1.79

McLuhan, M. (1964). *Understanding Media: The Extensions of Man.*

Odaci, H. (2011, August). Academic self-efficacy and academic procrastination as predictors of problematic internet use in university students. *Computers & Education*, *57*(1), 1109–1113. doi: doi:10.1016/j.compedu.2011.01.005

Olsen, F. (2001). Chapel Hill Seeks Best Role for Students' Laptops. *Chronicle of Higher Education*, *48*(4), A31. Retrieved from http://elibrary.ru/item.asp?id=4208553

Rigney, C. (2000, June). RADIUS Accounting. *Request For Comment*(2866). Retrieved from http://tools.ietf.org/search/rfc2866

Rigney, C., Willens, S., Rubens, A., & Simpson, W. (2000, June). Remote Authentication Dial In User Service (RADIUS). *Request For Comment*(2865). Retrieved from http://

tools.ietf.org/search/rfc2865

Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*(1), 41–55. doi: doi:10.1093/biomet/70.1.41

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics*, *121*(1-2), 297–317. doi: doi:10.1016/j.jeconom.2003.10.003

Rubin, D. B. (1973, March). Matching to Remove Bias in Observational Studies. *Biometrics*, *29*(1), 159–183.

Sacerdote, B. (2001, May). Peer effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics*, *116*(2), 681–704. doi: doi:10.1162/00335530151144131

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* (pp. 249–277). Elsevier B.V. doi: doi:10.1016/B978-0-444-53429-3.00004-1

Tindle, E. (2002). Pathological Internet Use (PIU) in University Students: A New Addiction. In *The 6th Pacific Rim First Year in Higher Education Conference Proceedings.* Christchurch, New Zealand: Queensland University of Technology.

Underwood, J. (2009, November). *The impact of digital technology.* Becta. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3059815

Underwood, J., Ault, A., Banyard, P., Bird, K., Dillon, G., Hayes, M., ... Twining, P. (2005, June). *The impact of broadband in schools.* Becta ICT Research. Retrieved from http://dera.ioe.ac.uk/1616/

Veiga, A., & Amaral, A. (2009, June). Survey on the implementation of the Bologna process in Portugal. *Higher Education*, *57*(1), 57–69. doi: doi:10.1007/s10734-008-9132-6

Young, J. (2008). Repression, Dissent, and the Onset of Civil War: States, Dissidents, and the Production of Violent Conflict [Paper 870]. *Electronic Thesis, Treatises and*

Dissertations. Retrieved from http://diginole.lib.fsu.edu/etd/870/

Zimmerman, D. J. (2003, February). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics*, *85*(1), 9–23. doi: doi:10.1162/003465303762687677