This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

Jump down to document ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

Browse Books & Publications
Make a charitable contribution

For More Information

Visit RAND at www.rand.org
Explore Pardee RAND Graduate School
View document details

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND PDFs to a non-RAND Web site is prohibited. RAND PDFs are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see RAND Permissions.
This product is part of the Pardee RAND Graduate School (PRGS) dissertation series. PRGS dissertations are produced by graduate fellows of the Pardee RAND Graduate School, the world’s leading producer of Ph.D.’s in policy analysis. The dissertation has been supervised, reviewed, and approved by the graduate fellow’s faculty committee.
Student Achievement, School Structure, and the Effects of Small Learning Community Implementation in Los Angeles

A Network Approach

Richard Bowman

This document was submitted as a dissertation in September 2010 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Paul Dreyer (Chair), Harold Green, and Francisco Martorell.
The Pardee RAND Graduate School dissertation series reproduces dissertations that have been approved by the student's dissertation committee.

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. RAND’s publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.
**Acknowledgments**

This production of this dissertation is the cumulative result of years of educational study and idea conception, and would not have been possible without the help of many other individuals. While I cannot mention everyone separately, I would like to take this space to acknowledge several particularly special individuals.

A special acknowledgement and thanks to my wife, Claudia, for providing perspective, love and support at all times, in ways I cannot adequately put into words.

A special acknowledgement to this dissertation’s funder, Douglas Smith, and the Susan Way-Smith Memorial fund, for their dedication to the cause and concept of a public education system; and to my committee members and external reader: Paul Dreyer, for his sage advice and support throughout the dissertation process and for being a most enthusiastic and interesting professor; Harold Green, for providing copious and most helpful comments on many dissertation drafts; Francisco Martorell, for being a great econometrics professor, and providing support throughout my PhD studies; and my external reader, Philip Kovacs, for his helpful comments on the draft document.

To my mentors, friends, and colleagues at the RAND Corporation and the Pardee RAND Graduate School: Jamie Gayton, for sharing his family, life advice, and leading the way, and whose warmth, friendship, and drive provided inspiration; Roberto Vargas, for a multitude of pep talks, strategic advice, and life lessons, and who helped me create and stick to a tight timeline; Benjamin Kim, who helped me stay focused and accountable during the writing
process; Yang Lu and Diana Epstein, for being great friends and mentors; and the current and former administrators, professors, and staff of the Pardee RAND Graduate School, for their warmth, dedication, can-do attitude, and whose overall vision for the school – to be the premier policy graduate school, interdisciplinary, practically focused, and quantitatively rigorous – provided the foundation upon which this integrative dissertation was built.

Finally, to my family, friends, former teachers and colleagues: Adriel Bowman, for being most logical and standing by me when the chips were down; Cecelia Bowman, for being a great example of perseverance by earning her doctorate a couple years before me; my brothers and sisters, for their support; Peggy Barber, for her knowledge, understanding, support, comments, hours spent editing, and great friendship; Deborah Steinert and Paul Del Rosario, for recognizing my potential many long years ago and pushing me to do something great (after I invented cold fusion and a cure for cancer, of course); Edward Victoria, for encouraging me to vigorously pursue what was right, regardless of the consequences; Katharine Paull, for her stubborn insistence that writing is a continuous process, not a product, and causing me to focus on my competitive advantage; Gloria Brewster, Marie Lara, and Stephanie Hancock from the California Institute of Technology for proving invaluable support during my undergraduate time there; and lastly to the teachers and staff of Sylmar High School and the Sylmar Math Science Technology Magnet, whose dedication to their craft provided me a wonderful secondary education, a rewarding experience as a teacher and colleague, and inspiration to help children.
# Table of Contents

Student Achievement, School Structure, and the Effects of Small Learning Community Implementation in Los Angeles ........................................................................................................................................ ii
A Network Approach .................................................................................................................................................................................. ii
Acknowledgments .................................................................................................................................................................................. iii
Table of Contents ................................................................................................................................................................................ vi
Tables .......................................................................................................................................................................................... ix
Equations ......................................................................................................................................................................................... xi
Figures ............................................................................................................................................................................................. xi
Abstract .......................................................................................................................................................................................... xiii

## Introduction
- Policy background ........................................................................................................................................................................... 1
- SLC benefits ..................................................................................................................................................................................... 1
- Types of SLCs .................................................................................................................................................................................... 2
- Intervention pathway ........................................................................................................................................................................ 3

## Review of Small Learning Community literature
- SLC studies ....................................................................................................................................................................................... 4
- SLC literature reviews .................................................................................................................................................................... 10

## Review of the school size literature
- School size studies ............................................................................................................................................................................... 11
- School size literature reviews .......................................................................................................................................................... 12

## Present research, aim, and roadmap
- Research questions ................................................................................................................................................................................ 14

## Chapter 1: Data Description and Methods
- Section 1.1: Sample frame and exclusions .................................................................................................................................. 18
- Section 1.2: Data description .......................................................................................................................................................... 18
- School data ...................................................................................................................................................................................... 19
- Data quality concerns ...................................................................................................................................................................... 20
- Section 1.3: General methods for student outcomes ................................................................................................................... 20
- Test scores ........................................................................................................................................................................................ 21
- Grade point average ........................................................................................................................................................................ 22
- Attendance ....................................................................................................................................................................................... 23
- Section 1.4: Network methods ...................................................................................................................................................... 23
- A network model of school structure ......................................................................................................................................... 23
- Measuring structure in a student course-taking network ........................................................................................................... 25

## Chapter 2: Community Structure
- Section 2.1: What is a community? ............................................................................................................................................. 27
- Section 2.2: Methods of community identification and evaluation .............................................................................................. 28
- Identifying communities by simplification .................................................................................................................................. 28
- Identifying communities by maximizing community quality ..................................................................................................... 30
- Algorithmic methods for community identification .................................................................................................................. 31
- Evaluating community similarity .................................................................................................................................................. 36
- Section 2.3: Community detection algorithm comparison results ............................................................................................... 41
- Quality of community structure .................................................................................................................................................... 41
- Similarity to known communities ................................................................................................................................................ 44
Tables

Table 1.1: Number of campus-grade combinations implementing SLCs by year ........................................ 20
Table 2.1: Examples of simplification options ............................................................................................ 28
Table 2.2: Concerns with different types of simplification ........................................................................... 30
Table 2.3: Similarity explanation .................................................................................................................. 37
Table 2.4: Predicted vs. implemented scalability of algorithms ................................................................. 49
Table 2.5: Predicted vs. implemented subjective scalability of algorithms ............................................... 49
Table 2.6: Community detection algorithm efficiency (higher is better) ................................................... 51
Table 3.1: Average school assortativity by student course choices ............................................................ 59
Table 3.2: Average school assortativity by student program participation ............................................... 60
Table 3.3: Average school assortativity by student academic outcomes .................................................. 61
Table 3.4: Average school assortativity by student attendance and discipline ......................................... 61
Table 3.5: Average school assortativity by student demographics ............................................................ 62
Table 3.6: Average school community segregation by student course choices ....................................... 66
Table 3.7: Average school community segregation by student program participation ............................ 66
Table 3.8: Average school community segregation by student academic outcomes ................................ 67
Table 3.9: Average school community segregation by student attendance and discipline ....................... 67
Table 3.10: Average school community segregation by student demographics ....................................... 68
Table 3.11: Variance decomposition of test scores, grades, and attendance .............................................. 73
Table 4.1: Sample extreme scenario ........................................................................................................... 78
Table 4.2: Structural descriptive statistics in 2004 .................................................................................... 82
Table 4.3: Structural descriptive statistics in 2008 ................................................................................... 83
Table 4.4: Structural model results ............................................................................................................ 83
Table 5.1: Correlations with percent of schools days attended ................................................................. 93
Table 5.2: Correlations with being suspended ......................................................................................... 95
Table 5.3: Correlations with GPA ............................................................................................................. 97
Table 5.4: Correlations with students having a high GPA (>3.3) ............................................................ 99
Table 5.5: Correlations with having a low GPA (<0.7) ............................................................................ 101
Table 5.6: Correlations with failing or dropping out ............................................................................... 103
Table 5.7: Correlations with reading test score levels .............................................................................. 105
Table 5.8: Summary of significant results ................................................................................................. 107
Table 6.1: Effect of SLCs on attendance ................................................................................................. 114
Table 6.2: Effect of SLCs on student suspensions .................................................................................... 115
Table 6.3: Effect of SLCs on grades ........................................................................................................ 117
Table 6.4: Effect of SLCs on dropouts/retention ..................................................................................... 119
Table 6.5: Effect of SLCs on test scores ................................................................................................. 121
Equations

Equation 3.1: Student level variance ............................................................................................. 71
Equation 3.2: Community level variance ...................................................................................... 71
Equation 3.3: Upper level variance ................................................................................................ 71
Equation 3.4: The community level ICC ...................................................................................... 72
Equation 4.1: Model specification 1: Base structural model ........................................................ 81
Equation 4.2: Model specification 2: Year controls ...................................................................... 82
Equation 4.3: Model specification 3: Size controls ...................................................................... 82
Equation 4.4: Model specification 4: Year and size controls ....................................................... 82
Equation 5.1: Model specification 1: Base model ........................................................................ 90
Equation 5.2: Model specification 2: Including SLC implementation ........................................ 91
Equation 6.1: Model specification 1: Base model ...................................................................... 112
Equation 6.2: Model specification 2: Additional covariates ........................................................ 113

Figures

Figure 2.1: Student subgroups in an SLC ..................................................................................... 40
Figure 2.2: SLCs inside larger student groups .............................................................................. 40
Figure 2.3: Comparison of community structure found (higher is better) .................................... 42
Figure 2.4: Comparison of community structure found using the weighted network (higher is better) ............................................................................................................................................ 43
Figure 2.5: Adjusted Rand similarity algorithm comparisons .................................................... 44
Figure 2.6: Jaccard similarity algorithm comparisons .................................................................. 46
Figure 2.7: Algorithm running times ........................................................................................... 50
Abstract

The division of Los Angeles’ large urban comprehensive high schools into groups of Small Learning Communities (SLCs) within the school campus was proposed as a way to improve academic outcomes. While the effects of school size on students have been explored in detail and converting school structure “in-place” is less costly than constructing several new small schools, little research has been completed regarding the structural or academic effects of dividing large schools into whole-school or “wall-to-wall” SLCs on the same campus.

With this policy and research backdrop, this dissertation defines and identifies communities of students, evaluates the level of sorting and segregation in schools and communities in schools, explores correlations between school structure and academic outcomes, and evaluates the effects of SLC implementation on school structure and academic outcomes. This analysis indicates that communities of students can be accurately detected algorithmically without prior knowledge of their number or size by using a novel social network analysis approach and defining student community as “students who take more classes with each other than with other students.” Further results suggest that, within schools, ability sorting is present; racial segregation is not present; and, extending Coleman (1966), the level of academic variation within student communities inside schools is at least as large as the variation between schools.

Structurally, whole-school SLC implementation caused a significant increase in the average number of classes shared with each other student in a child’s classes, and a significant decrease in the number of unique students in the average child’s classes. As suggested by sociological
research, children who shared more classes with each other also had better attendance, fewer
discipline problems, and higher GPAs; sharing classes with smaller numbers of students was
weakly protective against dropping out, with insignificant or complex correlations with other
outcomes. Finally, the implementation of whole-school SLCs significantly reduced suspensions,
dropouts, and the fraction of students earning very low GPAs, with no significant effect on
attendance or test scores after controlling for student characteristics.
Introduction

*Policy background*

In late 2004, the Board of Education (Board) of the Los Angeles Unified School District (LAUSD), the second largest school district in the nation, directed the superintendent to submit plans to convert all comprehensive high schools to a particular school structure: whole-school, or “wall-to-wall,” Small Learning Communities (SLCs). This school structure consists of reworking the school master schedule and students’ individual class schedules to form smaller groups of students (*communities*) who take classes together within the high school; simultaneously, groups of teachers who teach a smaller set of students are created. While groups of students who took classes together by design (e.g., magnet schools, teacher academies, Humanitas programs) or happenchance (e.g., combinations of course requirements and scheduled times that only allow one possible permutation) always existed, this was the first time that these communities were formed on a large scale using all students in a high school.

*SLC benefits*

In the Board’s request and in the superintendent’s policy document implementing the Board’s request, many attributes and effects were ascribed to the SLC school reform (Castillo, 2005). A partial list of these anticipated attributes:

- Have a distinct and compelling vision, mission, goals, objectives, and clearly recognizable identity
- Have between 350-500 students, and an inclusive admission policy
- Help close the achievement gap
- Provide coherent, rigorous standards-based curriculum
- Have an individual professional development component
- Cultivate community partners
• Focus on the needs of students within the SLC
• Create a spirit of collaboration
• Ensure equal access to post-secondary opportunities
• Increase the numbers of students graduating
• Provide a personalized learning environment where every child is well known by a group of educators

In approving this broad school reform, the LAUSD board echoed Capps and Maxwell (1999) when further writing that “educational research and experience demonstrate that students do better academically when they feel a connection to their schools, and small schools are a hopeful avenue for the improvement of urban school education” (Capps & Maxwell, 1999; Tokofsky & Lauritzen, 2004).

Types of SLCs
As of January 2008, the LAUSD had approved the SLC plans of fifty high schools, and had the plans of ten other schools under review, leaving only three existing high schools without SLC plans. As of May 2010, while more high schools have opened (as whole-school SLCs at inception), all high schools have approved plans and have implemented whole-school SLCs or are in the process of implementation. All of the plans include several SLCs in the form of career academies (Lerman, 1996), in accordance with the Board’s wishes (Flores Aguilar, Garcia, & Vladovic, 2008), with communities having thematic connections to students’ particular interests and future careers, potentially allowing students a greater level of personalization in their education compared to the vast expanse of the traditional comprehensive high school (DeJong & Locker, 2006). While many themed SLCs span all high school grades, some school plans also
allow for one or more 9th grade “houses” where groups of ninth graders are kept together before the transition to the career academy-style communities (Rayyes & Barela, 2009).

**Intervention pathway**

As exemplified in the language of the Board when approving this school reform, however, much of the research investigating effects of school size or “school smallness” relates to the effects of entire small schools (Raywid, 1998), or single SLCs, not the whole-school or “wall-to-wall” SLCs that this reform created (Allen & Steinberg, 2004). And while creating SLCs within schools may be one practical way to capture some of the effects of smaller schools without constructing new school buildings, they are relatively untested in the research literature.

**Review of Small Learning Community literature**

The theoretical underpinnings of the present research are most closely tied to research on high schools which have completely converted into small learning communities across their student body, reducing much of the bias from student or teacher selection into SLCs. However, specific research regarding academic outcomes and whole-school high school SLCs is quite lacking. Therefore, the universe of literature reviewed here is expanded to those on small learning communities in high schools, where communities are allowed to select students and teachers from a “host” school which must take the leftovers. Perhaps because SLCs are relatively new in the whole-school form, the available research is generally not plausibly causal, and often fails to address the selection on unobservables issue in the non-whole-school case – more motivated students or teachers might more likely choose to join the SLC if it is perceived as offering a better education, and other, less motivated students or teachers, may opt to stay in the “host school”, thus biasing the SLC intervention. To supplement these studies that are limited but more closely related to SLCs, after discussing available studies and then literature reviews on SLCs,
this review proceeds to an abbreviated recap of selected studies and then literature reviews of the research base regarding school size.

**SLC studies**
A large implementation study, (Bernstein, Millsap, Schimmenti, & Page, 2008), of the first cohort of 119 schools receiving SLC conversion funds from the US Department of Education found positive academic outcomes associated with SLC implementation. Specifically, the authors concluded that the percentage of students involved in extracurricular activities increased by five percentage points, that there was a positive trend in grade promotion rate from 9th grade to 10th grade, that the incidence of school violence dropped by 1.4 incidents per 100 students, and that the percentage of graduating students intending to attend two- or four-year colleges increased by four points. Additionally, there were downward trends in negative behavioral outcomes (disciplinary actions, school violence, and drug and alcohol use), and upward trends in graduation rates; however, early changes in scores on statewide assessments and college entrance exams after receiving SLC funding were mixed, modest, and non-significant. This study design did not include analysis of a set of comparable schools that would produce a reliable control group. In light of this, the authors suggest caution interpreting their findings, especially trends, because of the lack of data on other schools during the time period of study. Additionally, the target population included many schools which only implemented SLCs for some smaller population in the school, potentially resulting in attenuation of results (or no change at all, if one alternately presumes that the students most likely to benefit were able to select into the SLCs).

Further evidence of a positive association between SLCs and academic performance was presented by Lee and Friedrich (2007). The authors use data of 193 high schools receiving US
Department of Education SLC implementation grants in 30 states (M. Lee & Friedrich, 2007) along with a repeated measures ANalysis Of VAriance (ANOVA) to test whether the state percentile ranking of SLC grantees test scores changed over time in school years 2002, 2003, and 2004. After dropping 25 high schools due to missing rankings data, the ANOVA shows that the difference across years (within grantees) was significant, and a post-hoc analysis concluded that the mean difference in percentiles was small and positive with a mean effect size of 0.06, at most 2.4 percentile points. Further analysis of correlations between SLC attributes as provided in grant applications and student academic rankings provided inconsistent and mixed results.

Another comprehensive implementation study, (Shear et al., 2008), addressed the Bill & Melinda Gates Foundation’s National High Schools Initiative, which funded new small high schools and the conversion of large comprehensive high schools into separated, autonomous schools within high schools. The study found mixed or no results for these conversions. Specifically, for these conversion schools, it found that “personalization, high expectations, and respect and responsibility” increased while student engagement decreased, that quality of students’ work in English Language Arts (ELA) increased while the quality of mathematics work decreased, and that there was no change in rigor of instruction, relevance of instruction, attendance rates, grade progression rates, or test scores. As before, the study’s conclusions must be interpreted with some caution for several reasons. First, the study’s sample size is small, somewhat difficult to determine, and varies across outcomes. Specifically, the authors state they included “over 100 schools” in their data collection, but a closer look shows that each small learning community is counted as a separate school with that measure; after grouping SLCs into the originating comprehensive schools, it is clear that the results are based on no more than
twelve or seven original comprehensive high schools, with possibly as few as two SLC conversion schools in some cases. Additionally, the study was hampered by a lack of longitudinal data, sometimes drawing conclusions from one or two years of data, though the authors do note that results from case studies hint that the seven conversion schools surveyed had low prior levels of student achievement, were plagued by teacher staffing issues, and in general had many startup growing pains, which might contribute to finding mixed results from a short-term evaluation. Finally, the Gates Foundation conversion model ideal was strict autonomous SLCs (Shear, et al., 2008), a model that essentially involves breaking up a (semi-)coherent institution and setting off a power struggle for control of individual learning communities (Raywid, Schmerler, Phillips, & Smith, 2003), a process known to be fraught with peril (Muncey & McQuillan, 1996).

Another study, (James-Burdumy, Perez-Johnson, & Vartivarian, 2008), of twelve high schools that implemented ELA instruction reform followed by SLCs because of the Focus on High Schools (FOHS) initiative in Boston, found significant changes in student perceptions in three of twelve areas reviewed in their survey analysis: decreased student reports of misbehavior, increased use of “progressive” teaching methods, and diminished perceptions of their relationships with other students at the school. Their analysis of academic outcomes, using an interrupted time series model with twelve (seven before, five after) years of data, found significant beneficial changes in most academic outcomes studied: the number of days absent, number of days tardy, number of suspensions, number of unexcused absences, and percentage of students suspended all declined while the percentage of students promoted to the next grade increased. On the contrary, the number of days attended and test scores in ELA and mathematics
decreased. However, as the authors prominently state, their interrupted time series analysis over twelve years cannot be interpreted as causal because it is unable to separate trends over time from any change caused by the FOHS initiative. Their efforts to control for preexisting trends in outcomes using quadratic trend specifications over twelve years only serve to multiply this concern by raising additional questions about potential improper model specification (Britt, Kleck, & Bordua, 1996) – for instance, the author’s quadratic specification of the likelihood of suspension resulted in a prediction of a twenty percent suspension rate in the absence of the FOHS initiative, while historical data always ranged between four percent and seven percent. Furthermore, the interrupted time series model is sensitive to accurate information on implementation date (Glass, Willson, & Gottman, 2008). Though their research implicitly bases its interrupted time series analysis on a claim that all twelve high schools in Boston implemented SLC structure simultaneously in the 2003-2004 school year (James-Burdumy, et al., 2008), a prior SLC implementation study which included five of the Boston schools mentioned, (Allen, Almeida, & Steinberg, 2001), presents evidence in 2001 that definitively refutes this claim for those five schools, implies that all other high schools were at least in the process of changing structure by the 2000-2001 school year, and describes implementation or planned implementation of several temporally preexisting or co-occurring district-wide initiatives that would further discredit any causal interpretations. Still, while the results may not infer a causal relationship, James-Burdumy, et al. (2008) does hint that SLCs are associated with larger changes to behavioral outcomes than to test scores.

A study of a Baltimore high school reform initiative, (Smerdon & Cohen, 2009), corroborated the positive correlations between SLCs and better student outcomes. It used data from six newly
formed small “innovation” high schools and twelve whole-school SLCs formed from three of Baltimore’s seven comprehensive high schools. The authors concluded that students in the whole-school SLCs reported more positive safety, personalization, and administrative leadership environments than students in comprehensive schools; when lumped with the small innovation schools, they had higher test scores, attendance, and graduation rates. The study’s authors present evidence however, that selection on unobservable factors may have been a factor in these results: in some cases, the different SLCs formed from comprehensive schools had similar demographics, but very different academic performance. Additionally, the demographics of the “innovation” high schools were more academically advantaged than other schools. Smerdon & Cohen (2009) readily acknowledges these selection issues and presents conservative, non-causal conclusions.

Another small study of three high schools in Albuquerque which implemented whole-grade SLCs in 9th grade, (Johns, 2008a), and partial SLCs in their other grades, (Johns, 2008b), focused on “softer” academic outcomes from student and teacher surveys, excluding analysis of “long-term” outcomes like grades, passing rates, and test scores due to feedback from principals. This study found that teachers in freshman academies had a greater awareness of student needs and more unified expectations for student behavior and academics while freshman students simultaneously reported higher teacher academic expectations, peer support for academics, self-efficacy, and recognition for their efforts, compared to students prior to implementation; in the career academy, students reported a greater sense of belonging, more interesting and engaging classes, and more personalized attention from teachers.
A study of Chicago’s small schools initiative across all grade levels, (Wasley et al., 2000), yielded positive associations for students in small schools (including SLCs) and concluded that, compared to large schools, students in small schools had better attendance rates, much lower dropout rates, higher grade point averages, higher high school graduation rates, and increased reading achievement scores; violence occurred less frequently in small schools while parents participated more and were more satisfied with their schools. Unfortunately, the study included only three high school conversion “multi-schools” where the entire high school converted into an SLC format, with the vast majority of the small schools studied being separate SLCs in a “host” school. This limits the generalizability of their overall findings to whole-school SLCs: a closer inspection of their reported data shows that the three “multi-schools” performed poorly, with the overall positive results being driven by the other SLCs and small schools. This is probably to be expected; while Wasley, et al., (2000) included controls for observable characteristics, without some equitable limiting mechanism such as a lottery, the selection of students, families, and teachers into small schools or a SLC on a host campus inevitably included unobservable factors such as enthusiasm for learning (or teaching) which could plausibly be responsible for all of the reported positive associations for students in small schools.

Other original research on academic outcomes related specifically to the conversion of large comprehensive high schools into SLCs on the same campus is difficult to find. When present, it had very limited sample sizes (such as one school), or used quantitative methods (such as uncorrected Pearson correlations) with severe limitations, and is not discussed further in this review (Bemel, 2009; Cox, 2008; Smith, 2009; Sparger, 2005; Turnbo, 2007).
On balance, this review of original research on SLCs yields some evidence that students in SLCs tend to have better academic outcomes. Additionally, these correlations tend to be stronger for behavioral or emotional outcomes, with smaller or no correlations with test scores.

**SLC literature reviews**

There are a limited number of meta-studies on SLCs. The studies that exist provide insight into the gaps existing in the study of this reform. For example, in a literature review combining studies of school size, small schools, and small learning communities, (Raywid, 1996), lamented the lack of available large-scale research on downsized (via SLC or population) schools, but still concluded from the evidence that a number of these smaller schools had accomplished their stated purposes of better student attendance, more positive student attitudes toward school, greater academic productivity, and enhanced student and family satisfaction with school. However, no accommodation for selection on unobservables in the cited studies is made, and therefore it is impossible to credibly reject the hypothesis that higher performing students and teachers simply chose the smaller school or learning community (Matthews & Kitchen, 2007; Neild, 2005).

Another literature review, (D. Ready, Lee, & Welner, 2004), also lamented the lack of available research on whole-school SLCs, particularly on academic outcomes; concluded that the research that did exist supported a conclusion that whole-school SLCs improved school climate; and offered some cautionary evidence, using the de facto whole-school SLCs that are composed of different tracks inside year-round high schools, that SLC implementation might increase segregation and stratification inside high schools, especially under unrestrained student, family, and counselor choice (Delany, 1991; Frank et al., 2008).
These two literature reviews of research on SLCs provide more evidence that SLCs are associated with students with better outcomes. Though research was not plentiful, they repeat the refrain that SLCs tend to affect “softer” behavioral and emotional outcomes more so than test scores. Finally, they suggest that SLCs might increase students’ ability to self-segregate along academically motivated lines.

**Review of the school size literature**

The concept of small schools where “communities” of students learn together has existed in some form since 1896 (Dewey, 1896; Harms & DePencier, 1996). While many tend to link research on small (or smaller) schools with SLCs, several authors question whether research on school size and smaller schools can be validly and legitimately applied to inference about SLCs, (Allen, et al., 2001; V. E. Lee, Ready, & Welner, 2002). Still, to the extent that SLCs create an environment similar to a school of smaller size, that research remains relevant. Additionally, the research body regarding effects of school size on student academic outcomes is at a more advanced stage, as differences in school size have existed for as many years as there have been schools, allowing high-quality data to be collected for many schools across many different geographic locations. This has allowed researchers to assess the effect of different school sizes using rigorous methods on large datasets. As a sample of the results and methods used, two studies (one national, one statewide) and two recent comprehensive literature reviews of the literature on school size are discussed next.

**School size studies**

National longitudinal studies have found that smaller schools are associated with better academic outcomes, controlling for many other variables. While not directly addressing SLC reform, one
paper used the National Educational Longitudinal Study (NELS-88) and hierarchical models to
determine that schools which had “restructuring practices” were associated with higher student
gains in achievement and engagement (V. E. Lee & Smith, 1993, 1995). Additionally, higher and
more socially equitable engagement and achievement were consistently associated with smaller
high schools.

A recent rigorous study found positive effects for smaller school sizes at the elementary level
(Kuziemko, 2006). The author used Indiana state data, school mergers and openings (as
instruments for exogenous changes in enrollment), and difference-in-differences and 2-Stage
Least Squares (2SLS) models to determine the effects of a change in school size on math,
reading and attendance for several years after the change – concluding that smaller school size
caused increases in math performance, reading performance, and attendance.

**School size literature reviews**
The literature on school size is consistent throughout: smaller schools lead to better student
outcomes. The literature reviews by Cotton (1996, 2001) provide evidence of the overarching
consensus in the field about the theoretical sociological basis for expecting positive results from
SLCs. While conceding that inference on SLCs was limited, Cotton (1996, 2001) concluded that
school size literature generally provided evidence that in smaller schools academic achievement
is better. Furthermore, student attitudes toward school in general and subjects in particular are
more positive and students’ social behavior – as measured by truancy, discipline problems,
violence, theft, substance abuse, and gang behavior – was better. Other desirable outcomes, such
as student attendance in school, participation in extracurricular activities, and satisfaction with
those activities increased, and even student dropouts were lower. Finally, the author found that
students in small schools have better academic and general self-concepts, a greater sense of belonging, and better interpersonal relationships, among a host of other generally beneficial outcomes (Cotton, 1996a, 2001).

A large, recent literature review of high-quality research evidence going back to the 1960s on the effects of school size, (Leithwood & Jantzi, 2009), additionally citing and incorporating previous comprehensive literature reviews on the topic, (Andrews, Duncombe, & Yinger, 2002; Cotton, 1996b; Fox, 1981; Walberg & Walberg Iii, 1994), listed a multitude of effects of school size (including non-linear effects), and concluded that the weight of evidence clearly favored smaller schools, with students from disadvantaged backgrounds benefiting more from a smaller school than those students with less disadvantaged backgrounds – a conclusion that further suggests smaller schools might help reduce achievement gaps as well. In a comment incidental to its declared focus on school size and smaller schools, Leithwood & Jantzi (2009), also lament the lack of quality research on whole-school SLCs effect on academic outcomes, criticizes the meager research on potential pathways by which those SLCs may work, and concludes that whole-school SLCs have been much more talked about than studied systematically – a conclusion undisputed by this review.

While the research body of evidence on whole-school SLCs is in its infancy, research on the effects of school size provides solid theoretical and outcomes-based evidence that a smaller school size yields better outcomes for students. What research exists on whole-school SLCs tends to agree with a hypothesis that the SLC reform may produce some of the benefits of smaller schools by similar theoretical pathways (Levine, 2010).
Present research, aim, and roadmap

The present research focuses on one pathway by which the whole-school SLC reform can possibly improve children’s educational outcomes: reducing the effective size of each student’s cohort (simulating smallness) by changing course-taking structure so more students share more classes together. While the present research cannot separate any effects of SLCs on academic outcomes from any effects of changes in structure on academic outcomes, hypothetically, structural changes might diminish children’s feeling of being adrift in an ocean of other students, personalizing and enhancing the connection between a child and the school – a result relatively more important for disadvantaged or “at-risk” students (Cotton, 2001). This type of structural change, using modifications to student scheduling (Neubig, 2006), is inexpensive compared to other changes such as new school construction or school placement (Balfanz, 2006; Dewees, 1999). If not precluded by the typical inflexibility of schools’ master schedules (D. Ready, et al., 2004), it can provide a cost-effective aid to districts desiring to improve student outcomes.

The aim of this dissertation is to shed light on the link between school structure and academic outcomes and to investigate the changes in school structure and academic outcomes resulting from the implementation of whole-school Small Learning Communities (SLCs) in the Los Angeles Unified School District. To accomplish this aim, a novel application of social network analysis tools is presented and benchmarked, and five research questions are answered.

Research questions

RQ1: What is the pattern of performance across community detection algorithms?

RQ2: To what degree are students in schools sorted or segregated?
RQ3: To what extent did the implementation of SLCs change school structure?

RQ4: How are measures of school structure related to academic outcomes?

RQ5: To what extent did the implementation of SLCs change academic outcomes?

Chapter 1 begins with a discussion of general methods and concepts used in all chapters. This chapter also presents the high school as an affiliation network of students linked by the classes they share each day. Two other research teams have used similar “transcript study” methods to link students (Field, Frank, Schiller, Riegle-Crumb, & Muller, 2006; Heck, Price, & Thomas, 2004). Both Heck (2004) and Field (2006) used information on course names and levels (e.g., honors or remedial) to group students into types. Both Field (2006) and Heck (2004) studied one school, with small numbers of students, with 274 students analyzed in Heck (2004), and 149 transcript-units analyzed for Field (2006). This research builds on the basic linking concept, but vastly increases the size of analysis and by adding information about the time of day, teacher, and classroom, generates a much more detailed exposure network, and allows for grouping students into communities based on academic time spent together, not just type of classes taken.

To allow for comparison of student communities present before and after SLC implementation, Chapter 2 evaluates several different algorithms for detecting communities in schools without prior knowledge of structure (Research question 1). While Heck (2004) and Field (2006) have linked students together based on their shared classes and performed some type of computerized classification on students and classes, linking very strongly to literature on tracking (Friedkin & Thomas, 1997; Oakes & Guiton, 1995), and friend formation due to proximity (A. Gamoran, 1992; Hallinan & Smith, 1989; Hallinan & Williams, 1990; Kubitschek & Hallinan, 1998),
newer, more sophisticated algorithms for community detection have since been presented (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Clauset, Newman, & Moore, 2004; Diao, Li, Feng, Yin, & Pan, 2007; Du, Wang, & Wu, 2008; Du, Wu, Pei, Wang, & Xu, 2007; Duch & Arenas, 2005; Gulbahce & Lehmann, 2008; Karrer, Levina, & Newman, 2008; Medus, Acuna, & Dorso, 2005; Newman, 2006a; Newman & Girvan, 2004; Pons & Latapy, 2005; Raghavan, Albert, & Kumara, 2007; Reichardt & Bornholdt, 2004, 2006). This chapter presents the concept of a high quality community, adapted from Newman (2004) and as defined solely by the structure of the high school course-taking network. It further benchmarks the speed and accuracy of five modern, complex community detection algorithms on weighted and un-weighted representations of the student affiliation network (S. P. Borgatti & Halgin, 2010).

Several researchers have hypothesized that SLCs may increase sorting, segregation, or tracking in schools (Delany, 1991; Frank, et al., 2008; D. Ready, et al., 2004). Additionally, at least one study presented clear evidence that the implementation of SLCs on host school campuses was associated with student sorting (Wasley, et al., 2000). Chapter 3 builds on this research background by providing information on the level of sorting within schools in this school district using three different approaches (Research question 2).

The implementation of SLCs in Los Angeles should have a first-order effect on school structure by changing the classes that students take together (Price et al., 2003). Chapter 4 directly measures changes in school structure resulting from the implementation of SLCs, an approach that has not yet been taken in current research due to the small number of papers viewing the school as a course-taking network (Research question 3).
Small schools and SLCs are hypothesized to create their beneficial academic outcomes by such pathways as increasing personalization and less differentiation of instruction by ability (Cotton, 2001). Most studies of SLCs and small schools have found positive correlations between survey measures of personalization and SLCs. And there is some literature to support the hypothesis that changes in school structure can increase a sense of community (Royal & Rossi, 1996). While not causal, Chapter 5 adds to the literature by discussing the link between quantitative measures of school structure at the student level and academic outcomes, with the goal of providing background on possible pathways through which changes in school structure can affect academic outcomes (Research question 4).

Most studies have suggested that SLCs are associated with better academic outcomes (Cotton, 2001; Raywid, 1996), but less so for test-score outcomes (Bernstein, et al., 2008; James-Burdumy, et al., 2008; Shear, et al., 2008). Chapter 6 adds to this literature by presenting results from a large school district using a plausibly causal model, including changes in multiple academic outcomes associated with the implementation of SLCs (Research question 5). Finally, conclusions, a policy recommendation, and future research directions are provided.
Chapter 1: Data Description and Methods

The aim of this chapter is to provide a description of the data and general methods used throughout this analysis, including a presentation of the high school as a student course-taking network.

Section 1.1: Sample frame and exclusions

The data included all students enrolled in the Los Angeles Unified School district in grades nine through twelve in the school years 2003-2004 through 2007-2008. De-identified student identifiers were linked across school years to form a comprehensive dataset. This analysis includes students in traditional schools and magnet schools, excluding students in alternative, continuation, special education, non-traditional, or non-comprehensive types of schools. Students were considered to be present in the school where they received their final grades, as information on intra-year moves was not available. Students were excluded from the relevant analyses when they did not have the corresponding data (e.g., student missing test scores are excluded from test score analysis).

Section 1.2: Data description

The dataset included variables that describe student demographics and academic performance. Student demographic data included the following variables: gender, ethnicity, free or reduced price lunch status, English as a Second Language (ESL) status, special education status, and gifted status. Student attendance and discipline records were also included, as were student test scores on the state accountability tests. School structure information on each course taken by a student each semester and year was included: teacher identifier, room number, period, course
description, and marks for achievement and work habits. Finally, the dataset included the enrolling school name and address.

**School data**
Schools were grouped on the basis of their address (campus location), as students on the same campus can theoretically take classes together, though they may be in a different “school.” For example, a high school with an integrated magnet school may have two school codes, one for the magnet school, and one for the comprehensive high school; students can “cross over” and take classes in the other school, and this grouping allows for that circumstance. As can be seen from Table 1.1, as time passed, a few new schools were created.

**Small Learning Community data**
Each school implementing whole-school SLCs was required to produce a School Impact Report (SIR). These SIRs contain a detailed description of the SLCs a school proposed to implement, including the themes and classes, along with the timelines proposed for full implementation. SIRs were acquired for every high school implementing SLCs, and the date of wall-to-wall implementation for each grade level within the school was recorded. For example, if a school implemented “9th grade academies” before full implementation of “career academies” in all other grades, this was captured as a separate event.
Table 1.1: Number of campus-grade combinations implementing SLCs by year

<table>
<thead>
<tr>
<th>School year ending in</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of campuses</td>
<td>53</td>
<td>59</td>
<td>63</td>
<td>64</td>
<td>68</td>
</tr>
<tr>
<td>Count of schools</td>
<td>102</td>
<td>108</td>
<td>112</td>
<td>116</td>
<td>120</td>
</tr>
<tr>
<td>Number of campus-grades implementing SLCs</td>
<td>3</td>
<td>12</td>
<td>55</td>
<td>96</td>
<td>142</td>
</tr>
<tr>
<td>Number of campus-grades</td>
<td>214</td>
<td>224</td>
<td>238</td>
<td>253</td>
<td>267</td>
</tr>
<tr>
<td>Percentage of campus-grades implementing SLCs</td>
<td>1.4%</td>
<td>5.4%</td>
<td>23.1%</td>
<td>37.9%</td>
<td>53.2%</td>
</tr>
</tbody>
</table>

Data quality concerns

The primary data concern is that the date of wall-to-wall SLC implementation was determined from the School Impact Report (SLC plan) each school filed. This presents an identification concern if the school did not implement SLCs as planned (James-Burdumy, et al., 2008).

Although a full census of schools was not completed, anecdotal evidence from informal interviews with teachers and administrators suggests that schools did implement SLCs in accordance with the timeline in their finalized and filed plans.

Another data concern is that learning communities, such as magnet schools or Humanitas programs, existed in schools before the school formed whole-school SLCs. These are some of the communities previously analyzed with favorable results – though ignoring selection bias (Raywid, 1996). This analysis does not attempt to account for partial implementation of learning communities before the school formed whole-school SLCs. Therefore any observed effects of SLCs should be viewed as conservative with regard to effects on the average student.

Section 1.3: General methods for student outcomes

This section describes the student outcomes of interest. The literature on student outcomes, small schools and SLCs show that both academic and behavioral factors may be affected by school structure. This research focuses on five domains: test scores, grades, attendance, discipline, and grade retention.
Test scores
To allow for comparison of the state accountability test (California Standards Test) scores across years and grade levels, test scores were converted from scale scores to z-scores. With this conversion, effect sizes are interpreted as fractions of a standard deviation (e.g., an effect size of 0.5 implies an effect of one-half of a standard deviation). To do this, the empirical scaled test score distribution in the school district was inverted, and then each scaled test score was assigned a score from the corresponding normal distribution with mean zero, and standard deviation of one (Langoff, 1996). Test scores were absent for students in the twelfth grade. Test scores were normalized within district by year, grade, and test.

While this transformation (into an interval scale) allows for a similar interpretation across years and grades (i.e., as positions in the standard normal distribution), it implicitly compares each student to other students in their own grade. This, and the fact that students who proceed to higher grades (i.e., that do not fail or drop out) score higher on average, yields the result that a student’s z-score across grade levels (over time) will fall for a student whose absolute performance (and implied ability) is constant. This is because the student’s relative position (i.e., z-score) is being calculated from a cohort with increasing average test scores (and implied ability) across grades. In the analyses to follow, this is controlled for by calculating SLC test score effects relative to non-SLC test score effects. In other words, the average drop in z-score across each grade level is effectively subtracted before calculating results.

In California, the math, social studies, and science tests are not given each year to every student, and the particular math and science test taken is chosen by the school and students through their class selection. This causes complications when attempting to interpret those normalized z-
scores, as the z-score conversion is conditional on the student population taking the test. Furthermore, schools can choose to give an easier test to their students by changing the grade levels when students take courses of different types. Accepting this situation and avoiding these limitations, this analysis generally uses the English test, which is given in each tested year, grades 9-11, and is the same for all students in a particular grade level in a particular year.

**Grade point average**
Student annual grade point average (GPA) was calculated by converting letter grades into a five point scale, where an “A” corresponds to a “4” and a “Fail” corresponds to “0.” The number grade was weighted by the number of credits that grade was worth. The annual grade point average is used for maximal inclusion of students, and further to avoid the problems inherent in calculating and comparing cumulative GPA when all grades are not observed for all students.

There is one limitation on the interpretation of GPA correlations and effects that is particularly relevant for this research. The structural portion of SLC reform can change GPA without influencing student learning if SLCs change the average level of ability grouping and teachers grade students relatively (i.e., on a curve) to any degree, and that degree is not constant across performance levels (i.e., if teachers are not indifferent between assigning all their students As or Fails). Evidence dating back to 1912 indicates that this was the case then, and is still the case (McMillan, 2001; Starch & Elliott, 1912, 1913a). By changing class structure, the SLC reform might change the population of students inside any given teacher’s classroom or overall classes, and particularly the degree to which higher level and lower level students share the same classroom (or teacher). If SLC reform increases the amount of this ability-based grouping, and higher level students receive good GPAs regardless of how many other high level students are in
their class (e.g., teacher assigns all As), then average GPA will increase because some lower level students will get higher GPAs than otherwise would be the case (e.g., teacher assigns some students Cs or Ds for effort, instead of Fails). This teacher grading scale effect is separate but cannot be isolated from differences in actual learning resulting from coincidental factors such as the hypothesis that the teacher may be able to teach more effectively to a more hetero- or homogeneous ability group (Argys, Rees, & Brewer, 1996; Wright, Horn, & Sanders, 1997), or the “big-fish-little-pond effect” (BFLPE) where some students in the low level class try harder (learning more) because they are still the best students in the class (Alwin & Otto, 1977; Marsh, 1987; Marsh & Parker, 1984).

**Attendance**
Student attendance was calculated as a percentage of days enrolled which were actually attended. Additionally, students were flagged as habitually truant if they were absent more than 20% of the time (regardless of reason).

**Section 1.4: Network methods**
A network model of school structure
To measure school structure quantitatively, an approach from social network analysis and graph theory is used on the student course-taking network (Field, et al., 2006; Heck, et al., 2004). This approach allows a quantitative measurement of school structure. Each school is considered to be a 2-mode network. The combination of a specific course, classroom, teacher, and time of day is an event, which is referred to as a teacher-class-time triad, or simply class, and the student is an actor. Connections are present between students and the classes they were enrolled in at the end of the semester. In graph form, this is an undirected graph, both students and classes are vertices, and edges are present when students attend a class; in matrix form, this is a rectangular binary
adjacency matrix where students correspond to rows, and classes correspond to columns. This representation is the 2-mode student-class network, and it is the most complete representation of the data available because it contains all the information available. However, this analysis uses a simpler network representing only how students are connected to each other, ignoring the duality of students and classes, unlike Field (2006), because of the greater research support base for the simpler networks.

The type of network mentioned above is called an affiliation network, and can be formed from the matrix representation of the 2-mode student-class network by multiplying the binary adjacency matrix by its transpose. The resulting matrix has dimensions equal to the number of students; the entries of the matrix represent the number of classes the two students in that row and column share, with the diagonal having the special interpretation of the number of classes each student takes. These correspond to undirected graphs where students are vertices, and edges are present and weighted by the number of class-teacher-time events that are shared between the vertices. This representation is the 1-mode weighted student affiliation network. The final representation simplifies these 1-mode networks by eliminating edge weights. The interpretation in this case is that students are connected if they share at least one teacher-class-time triad. This representation is referred to as the 1-mode un-weighted student affiliation network, and can be visualized in 2-dimensional space using a force-directed layout (Fruchterman & Reingold, 1991; Ihaka & Gentleman, 1996).

Separate networks for each combination of school location, year-round track, grade level, year, and semester are used for community detection (i.e., communities are calculated within grade
and year-round track) to be able to accurately find communities (Fortunato & Barthélemy, 2007). The entire network of students in a school location in a given year and semester regardless of their track or grade level is used to calculate school structural measures.

**Measuring structure in a student course-taking network**
Centrality in the context of students in a high school course-taking affiliation network is a way of quantitatively describing each student’s daily school experience. In other words, does the student see the same set of 30 students throughout the day – a member of a close-knit cohort, or is the student exposed to 400 different students per day – bobbing alone in an ocean of people as in Cotton (2001)? Freeman’s (1977) paper outlines a type of centrality, degree, that is used here in a novel application to students’ school course-taking affiliation networks (Freeman, 1977).

**Calculating degree and structural measures**
In the student course-taking affiliation network context, a student’s degree (D) is the number of other students he or she shares classes with. In network (graph) terms, it is the number of other students (vertices) connected to that student (vertex). Mathematically, it is the row sum of the 1-mode un-weighted student affiliation network matrix, and is calculated by the following formula,

$$ D(v) = \sum_{p \in V} A_{vp} $$

where v and p are vertices representing students, and A is the un-weighted binary adjacency matrix (Stephen P. Borgatti & Everett, 2006; Wasserman & Faust, 1994). A student with a smaller degree is exposed to a smaller number of students each day, and a student with a larger degree is exposed to a larger number of students each day. This can be viewed as a student-centric measure of school size, and according to literature, should be correlated with better academic outcomes (Leithwood & Jantzi, 2009), and decrease when SLCs are implemented (Cotton, 2001).
Strength (S), or weighted degree, is similarly calculated, but uses the weighted adjacency matrix instead of the un-weighted adjacency matrix. Mathematically, 

\[ S(v) = \sum_{p \neq v} A^w_{vp}, \]

where \( v \) and \( p \) are vertices representing students, and \( A^w \) is the weighted adjacency matrix. In this case, the weights are the number of classes shared; the weighted degree divided by the degree yields the average number of classes (C) shared with each connected student: 

\[ C(v) = S(v)/D(v). \]

This measure can be viewed as an estimate of the level of cohort structure, and should be positively correlated with better academic outcomes (Raywid, 1996), and increase with SLC implementation (Cotton, 2001). The analysis using these three measures (degree, strength, and the average number of classes taken with each other connected student) excludes students with a degree less than or equal to one (i.e., any student who is only connected to one other student).
Chapter 2: Community Structure

Section 2.1: What is a community?

What is a community in a high school context? The authorizing LAUSD board motion and resulting concept paper mentioned many attributes of a SLC that might be viewed as definitions of community (Price, et al., 2003). But, is community a contiguous location on campus? Is it a sense of shared identity among students? Is it a group of teachers who agree to collaborate together? Is it simply an administrative label? Or is it students who take classes with each other? While all of these things are useful concepts when thinking of “community” in different contexts, the structural focus of the present research leads to a focus on the last concept of community: a student community defined as those students who take more classes together than with other students. A small school implicitly creates this type of community, and SLCs should also replicate it, if unhindered by the master schedule (Neubig, 2006). This is a structural definition of community. While students who share more classes together might have spontaneous community feelings (Bryk & Driscoll, 1988; Royal & Rossi, 1996), those are separate from this definition, and not addressed directly with this research.

Defining a community as students who take classes together is intuitive, but quite difficult to quantify. A large urban comprehensive high school offers many combinations of elective classes and core classes, and students themselves have varied interests and are generally allowed to pursue electives and courses of their choice, subject to meeting prerequisites. This large menu of courses and ability of students to choose based on natural variation in student interests mean that some students in any group of students designated as a community will also take classes with other students not in the community. The research that comes closest to this approach uses a less
sophisticated clustering method, only one high school, and reduced student transcripts across all years to show that “emergent tracks” of ability sorted students can form without explicit tracking actions by administration (Heck, et al., 2004; Oakes & Guiton, 1995).

Section 2.2: Methods of community identification and evaluation

Identifying communities by simplification

One easy way to address the consequences of students’ ability, in the comprehensive high school (Powell, Farrar, & Cohen, 1985), to take a wide variety of classes with a wide variety of students and still identify groups of student communities from their classes is to eliminate some subset of classes (e.g., all non-academic classes) and form the communities by grouping students using the other classes (e.g., all student who share at least two academic classes). This solution takes the following general form: a community is formed by students that share: some number or fraction of some set of class types classes together. This presents a two-dimensional criteria scale: the first dimension being inclusivity (the number or fraction of classes required to form a link), and the second being representation (the types of classes included in the link). Choosing where to locate analysis on this two-dimensional scale requires essentially arbitrary judgments that are difficult to defend, or have unacceptable tradeoffs as explored separately below.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Inclusivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strong</td>
</tr>
<tr>
<td>Strong</td>
<td>Share any class</td>
</tr>
<tr>
<td>Weak</td>
<td>Share English class</td>
</tr>
<tr>
<td></td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td>Share all classes</td>
</tr>
<tr>
<td></td>
<td>Share all English classes</td>
</tr>
</tbody>
</table>

Table 2.1: Examples of simplification options
**Inclusivity**
Decisions surrounding what number or fraction of classes to use require choosing how inclusive communities are. Because it is rare that students have identical schedules (Field, et al., 2006; Heck, et al., 2004), forming communities based on a requirement that all classes are shared with all students in that community eliminates many weaker connections (e.g., students that share all classes except one) that may also be important, and produces an unacceptable number of isolated students with no community identification because they do not share an exact schedule with anyone else. Relaxing the number or percentage of shared classes required to form a community reduces the number of isolated students, but can quickly and erroneously imply that all students are in the same community.

**Representation**
Decisions about what types of classes to include also require arbitrary or inflexible decisions. Decisions on this dimension represent a tradeoff on how representative the resulting community is to actual communities. The maximally representative choice, where all classes are included when forming communities, when combined with a weak inclusivity criterion, places all students together, for example, forming a community by students sharing at least one (weak inclusivity) of any class; and generates an unacceptable amount of isolates with a strong inclusivity criterion, for example, forming a community by student sharing all (strong inclusivity) classes. A minimally representative choice, choosing communities formed by one class, creates a number of easily defined communities (i.e., each class is a community), but they may not be clearly related to which students actually spend most of their time together (e.g., forming communities based solely on math class omits students who share English and history, but take different math classes).
### Table 2.2: Concerns with different types of simplification

<table>
<thead>
<tr>
<th>Representation</th>
<th>Inclusivity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Strong</strong></td>
<td><strong>Weak</strong></td>
</tr>
<tr>
<td><strong>Strong</strong></td>
<td>Only one community</td>
<td>Too many isolated students</td>
</tr>
<tr>
<td><strong>Weak</strong></td>
<td>Less representative and/or only one community</td>
<td>Less representative and/or too many isolates</td>
</tr>
</tbody>
</table>

Community formation by simplification is inherently and unacceptably inflexible and crude. Not only does it require arbitrary choices about which and how many classes to include or exclude, it necessarily excludes potentially valuable information about course structure.

**Identifying communities by maximizing community quality**

*Community quality: Modularity*

A better method for detecting communities is to determine a maximally inclusive and representative community quality function. Then select the grouping of all students that has the highest quality as measured by this function. This method is conceptually elegant. Depending on the quality function, it can also use all information about students and provide a relative measure of how strong community structure is in a school.

A good definition of community quality would measure the extent to which students inside each community take relatively more classes with other members of the community than with those outside that community, compared to those that would be expected by chance. A group of researchers recently presented such a measure (Newman, 2004). Called modularity, and by convention denoted with a $Q$, it represents exactly this definition of community quality. In equation form, $Q = \sum_i (e_{ii} - a_i^2)$, where $e_{ii}$ represents the fraction of all links that connect within community $i$ (i.e., fraction of all links between students that are between students in community...
$a_i$ represents the fraction of total links that connect to community $i$ regardless of whether they connect to other communities (i.e., fraction of all links in the school which connect to any student in community $i$). Modularity has a range from negative one to one: negative values imply that students tend to take more classes with students outside of their designated communities; a value of zero implies that the distribution of classes taken is equivalent to chance; positive values indicate students tend to take more classes in their groups than could be expected by chance. While positive values indicate a community grouping where students take more classes within their group than outside their group, the developers of this measure experimentally determined that real-world networks with “significant” community structure started to appear at modularity values starting from about 0.3 (Newman & Girvan, 2004).

**Algorithmic methods for community identification**

**General methods**
Having determined a community quality function, we now have to simply pick a grouping of students that maximizes this quality function. However, there are too many possible groupings to naively evaluate each possible grouping for a maximum individually; the number of possible groupings in a set of size $n$ is given by the $n$th Bell number, which increases exponentially (Rota, 1964). As an example, for a set of 500 students, there are about $1.6 \times 10^{843}$ possible groupings. Therefore heuristic approaches, which cannot ensure that an absolute maximum quality is found, must be taken. There are two basic types of heuristics: agglomerative approaches, which build communities piece by piece, and divisive approaches, which start with one community and split it into pieces. This chapter evaluates five different unassisted algorithms which use different approaches to find communities. Generally, however, each algorithm proceeds by transforming the student-to-student links (shared classes) in a school into another form (e.g., a map of
distances between cities), calculating groups in that other form (e.g., finding regions that include cities that are close together), and translating the results from those calculations back into student communities (e.g., labeling each student with their proper region, or community). Finally, the quality function, modularity, evaluates the level of community structure found by each approach.

**Label Propagation method**
The label propagation (LP) method is the simplest and quickest algorithm (Raghavan, et al., 2007). It is an agglomerative approach where each student is given a random community label and, at each step, each student takes the label of the majority of his or her connected peers, terminating when no more changes can be made (i.e., each student has the label of the majority of his or her connected peers). At its conclusion, students with identical labels form the communities, and the community quality is measured with the modularity function.

**Random Walk-trap method**
The random walk-trap (WT) method of community identification is an agglomerative approach which relies on the fact that following random links between students who are in a community tends to yield other students who are in the same community (Pons & Latapy, 2005). It is described as follows: first treat each student as their own community. Then, for each community, calculate the probability that each other connected community should be merged with that community by sequentially following a set number of random links between communities. Merge the two most likely connected communities, save the group memberships, and repeat the process. The result of this algorithm is a list of the merges and community memberships in order of occurrence. From this list, the modularity of all community groupings is calculated and the student community grouping with the highest modularity is chosen as the final result.
**Fast Greedy modularity optimization method**
The fast greedy (FG) method of community identification directly optimizes modularity by an agglomerative method, finding a local maximum of community quality (Clauset, et al., 2004).

This algorithm proceeds as follows: first begin with each student in their own community. Then for each community, calculate the increase in modularity formed by joining each other connected community. Then, merge the two communities that would give the greatest increase in modularity, and repeat the process. The result of this algorithm is a list of the merges and memberships in the order they occurred. From this list, calculate the modularity of the student community membership for each merge and pick the community grouping with the highest modularity.

**Leading Eigenvector method**
The leading eigenvector (LE) method of community identification is a divisive method, recursively splitting schools into separate communities of students (Newman, 2006b). This algorithm proceeds as follows: first transform the students’ links into one adjacency matrix. Then form a “modularity” matrix by subtracting a matrix representing the expected number of links between each student given their respective degree measures (assuming links were distributed randomly across all students in the school network) from that adjacency matrix. Then calculate the eigenvector corresponding to the largest positive eigenvalue of this modularity matrix, and then split students into two groups based on the signs of the elements of this eigenvector. Repeat this process on each piece (with modularity calculated from the original modularity matrix) until there are no more positive eigenvalues of the adjacency matrix. At this point, the groups formed by all of the splits are the student communities. The quality of this set of communities is then calculated using modularity.
**Potts Spin Glass model method**
The Potts spin glass model (SG) method of community identification is also a divisive method, but instead of recursive splitting, it forms all divisions simultaneously (Reichardt & Bornholdt, 2006). Reichardt and Bornholdt (2006) showed that the process of maximizing modularity was equivalent to finding the ground state (i.e., minimizing the energy) of an infinite range Potts spin glass, an atomic-level quantum model of generalized ferromagnetism and anti-ferromagnetism in a frustrated crystalline lattice usually studied by condensed matter physicists. Interpreting the school as this spin glass model, each student is considered to be a single atom in the spin glass lattice, which may or may not interact with other atoms in the spin glass; students’ community identification is considered to be one of many different types of spins (i.e., magnetic states) that each atom can have. Student links (i.e., atom interactions) across communities (i.e., to atoms with different spins) have higher energy, and student links within communities (i.e., with the same spin) have lower energy. The sum of all the energy across all the atomic interactions (i.e., student links) in the spin glass is the called the Hamiltonian energy function for the model, which must be minimized to find the ground state and therefore the community grouping with maximum modularity. This minimization is performed using a linear simulated annealing Metropolis algorithm – a heuristic method that uses randomization to improve results (Kirkpatrick, 1984). To minimize the Hamiltonian energy function, each student’s community identification (i.e., spin) is successively changed to the community that reduces the overall energy (i.e., the Hamiltonian) the most by calculating the change in energy resulting from each student (i.e., atom) changing from their current community to each other possible community. However, since the Hamiltonian energy function may have several local minima, to avoid getting stuck on one of the local minima and finding a non-optimal community division, at random, student community changes that increase total energy less than a limit are allowed. Over time, as
more students’ community identifiers are changed, the allowable energy increase for these random changes is slowly reduced until the Hamiltonian energy function reaches a minimum (and modularity therefore reaches a maximum) and no allowable change can be made (i.e., there is no energy decreasing change, and any random change would increase energy more than the allowed limit). At conclusion, the spins of the atoms representing each student are their corresponding community identification, and the quality of the communities thus formed is measured using modularity.

**Limitations**

All these algorithms share the limitation of using a community quality function, modularity (Karrer, et al., 2008). At algorithm conclusion, there is no way to easily determine the internal validity of groupings, other than that modularity-based quality function (Hastings, 1989). This section addresses internal validity by adding a quality measure based on similarity to a known grouping.

These community detection algorithms can also be classified into deterministic algorithms and non-deterministic algorithms. LP, LE, WT, and FG are deterministic algorithms, which find the same result regardless of how many times they run, like the algorithms used in Heck (2004) and Field (2006). These deterministic algorithms share the weakness that they can only find one possible grouping. Because there are so many possible groupings of students, it is unlikely that they can find the absolute best grouping.

SG is non-deterministic, its results have a random quality, and can possibly change each time it is run; still, while it finds a minimal energy level, there is no assurance that it is the absolute
minimum energy and, while it finds a community grouping associated with that minimal energy, there is no assurance that it is the unique community grouping with that energy. Still, because it tries so many combinations, it can provide better results than the single-shot deterministic heuristic methods.

Finally, all groupings given by these algorithms are unlabeled. They do not have names such as the “Math Learning Community” or “Business Learning Community.” However, this limitation is sidestepped because this analysis is more focused on the structure of schools, and not the particular content (i.e., the typical courses taken by the students) of each community.

**Evaluating community similarity**

The purpose of this section is to measure how well the communities found by the community detection algorithms match actual known SLCs. Because the algorithms use a community quality function, modularity, that uses structural information only, and does not consider the type of classes in its determination of quality, it might be possible that the communities are unrelated to the actual SLCs present in a school. To address this threat to internal validity and to determine how well the algorithm results match known SLCs, they are compared to a known SLC code when present in the data set (Hastings, 1989). While there are many comparison measures available to compare different community groupings, for example, (Danon, Díaz-Guilera, Duch, & Arenas, 2005; Dimitriadou, Weingessel, & Hornik, 2002; Fowlkes & Mallows, 1983; Katz, 1953; Milligan & Cooper, 1986; Strehl & Ghosh, 2003) – because most of them give very similar results for the same groupings, as recommended by a review of several measures, two conceptually simple methods are used to compare similarity here (Milligan & Cooper, 1986) – Jaccard similarity and Rand similarity adjusted for agreement by chance.
**Jaccard similarity**
The first method, Jaccard similarity, was originally used to compare the distribution of different species in different locations across the Alpine mountains (Jaccard, 1912). Its extension to student communities is as follows: Jaccard similarity is the ratio of the number of pairs of students classified together in both the known community and the algorithm to the total number of pairs of students classified together in either the known community classification or the algorithm classification. Jaccard similarity is \( J = \frac{a}{a + b + c} \), where the letter meanings are as indicated in Table 2.3. Jaccard similarity ranges between zero and one; a value of one represents identical community classifications, a value of zero represents no agreement at all.

<table>
<thead>
<tr>
<th>Table 2.3: Similarity explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of student pairs</strong></td>
</tr>
<tr>
<td>Algorithm-classified together</td>
</tr>
<tr>
<td>Correctly together</td>
</tr>
<tr>
<td>Algorithm-classified apart</td>
</tr>
<tr>
<td>Incorrectly apart</td>
</tr>
</tbody>
</table>

**Rand similarity**
The second type of similarity, Rand similarity, reflects the idea that having two students separated should be correctly separated is as important as having those two students correctly grouped together (Rand, 1971). Rand similarity is: \( R = \frac{a + d}{a + b + c + d} \), where the letters have meanings as indicated in Table 2.3. Because two random sets of communities are very likely to have students correctly classified both apart and together, this analysis uses a version of this Rand index which is adjusted for being correctly classified by chance is used from (Hubert & Arabie, 1985), and defined by \( R = \frac{a + d - n_c}{a + b + c + d - n_c} \), where the adjustment for chance is based on the overall number of students, \( n \), the number of students in each community calculated by the
algorithm, \( n_i \), and known previously, \( n_j \), in the following way:

\[
    n_c = [n(n^2 + 1) - (n + 1) \sum_i n_i^2 - (n + 1) \sum_j n_j^2 + 2 \sum_i \sum_j n_i^2 n_j^2 / n] / [2(n-1)].
\]

The corrected Rand similarity ranges between zero and one, where one represents identical groupings, and zero represents no more agreement in groupings than would be expected by chance.

**Data quality concerns**

The primary data quality concern is the precise identification of student community membership.

The known SLC code variable is entered by hand, and is subject to two general data quality concerns. The first general area of concern is missing data: first, this variable is uncollected and therefore missing before the 2005-2006 school year; second, even after the 2004-2005 school year, schools sometimes did not use the administrative variable to track student SLC, instead choosing to track students with another locally created variable not included in this dataset; third, this variable is often partially missing by school, even where schools have SLCs.

The second general area of concern is inaccurate data; occasionally the known SLC variable appears to be used for some other purpose entirely, not corresponding to any coherent pattern of classes after a hand inspection of classes taken and descriptions of actual SLCs present. Second, when a student transfers schools, sometimes the SLC code from the previous school is included even though it is not present in the current school. Third, since the variable is entered by hand, mistakes in data entry are always possible. Finally, sometimes an SLC code is present for everyone, even when full SLCs are not present in that grade; because these codes do not appear to match anything, perhaps they reflect SLC choices in the following year.
Additionally, the meaning of all the SLC codes used was not completely clear. Because this analysis does not rely on the type of SLC or community, however, precise identification of SLC name is unnecessary, only that students that share the same code share the same community. Still, there seemed to be some evidence that different SLC codes occasionally corresponded to students in the same SLC, although determining this precisely was impractical. Because this analysis is simply used to rate algorithms, and all algorithmic methods would be similarly biased by this issue, it has limited effect.

**Competing groupings**
The hypothesis regarding similarity is that among the community detection algorithms, one should be the best, but not necessarily match exactly. It might not be reasonable to expect a community detection algorithm to match the known SLCs exactly (Gulbahce & Lehmann, 2008). Several scenarios could lead to this outcome. First, since data from courses taken is more thoroughly cross-checked (since it affects graduation) than the SLC code, and additionally because these algorithms aggregate information across so many course records, small errors should not make as great a difference in the final placement of students into communities. An error in SLC code, on the other hand, directly causes an error in community classifications.

Second, the groups formed by the algorithm might be valid subgroups of the known SLC, as shown in Figure 2.1.
Similarly, known SLCs might be valid subgroups of a community found algorithmically, as shown in Figure 2.2.
While a perfect match is desirable, algorithms that find smaller or larger communities might be acceptable (Gulbahce & Lehmann, 2008; Gustafsson, Hörnquist, & Lombardi, 2006). SLCs try to group students, but there are other, competing groupings. For example, schools may change advanced classes into “passport” classes, where a student in any SLC may take them (Price, et al., 2003). All advanced students may then be grouped together if they all take a significant number of these courses. Similarly, if there is a set of students that need remediation, or are interested in some particular subject, the degree to which those classes are shared by students across all SLCs creates a competing community that may be found by a good community detection algorithm, but not match the administrative SLC.

Section 2.3: Community detection algorithm comparison results

Quality of community structure
The modularity of the best community structure from each algorithm was calculated in addition to the modularity associated with the communities formed by the SLC code. Because SLC code is not always present, comparisons were limited to the portion of schools and grades that had completed SLC codes for each student (regardless of whether those codes corresponded to a detectable community structure). The data was then binned by deciles of community quality modularity associated with the SLC code, and the average for each bin graphed for each of the community types. This grouping shows how well the community detection algorithms are doing when the SLC code contains a high level of community structure, as well as when it does not. The results are displayed in Figure 2.3.
As can be seen from Figure 2.3, while all approaches except the LP approach are quite correlated, the SG approach consistently has a higher modularity score over all deciles of SLC modularity. In addition, the fact that the graph of modularity has an upward sloping trend with SLC modularity suggests that both are measuring similar attributes in school community structure. When the SLC code indicates relatively little community structure, all approaches find relatively little community structure. When the SLC code indicates a high level of community structure, all approaches tend to find a similar level of structure. The divergence in the pattern of modularity below the median SLC modularity is of note: this pattern indicates that the SLCs given by the SLC variable in those schools may be SLCs in name only, with their students sharing not many more classes than those expected by chance. The LP approach fails to find any
community structure until the actual structure is very high, at which point it works about as well as the others.

A similar pattern is shown in Figure 2.4, using a weighted modularity, where links between students are weighted by the number of classes shared, to measure quality of community structure found. The weighted SG approach again finds the highest community structure, while FG, WT, and LP follow, in that order. Again, below the median modularity, the SLC code modularity diverges from all other approaches. Overall, while the best weighted modularity levels found are higher than un-weighted modularity levels, the two are not directly comparable.

Figure 2.4: Comparison of community structure found using the weighted network (higher is better)
Similarity to known communities

*Adjusted Rand similarity*

The similarity of each algorithm with respect to the SLCs was calculated using the adjusted Rand similarity, as well as the similarity of a null grouping (NUL) where every student was in the same community. From Figure 2.5, it is clear that of the different approaches and weightings, the SG algorithm has a clearly higher similarity to the known community structure, with the similarity increasing as SLC modularity increases, and correctly classifying students together and apart about 88% on the time. Again, it can be seen that as modularity increases, all the unassisted approaches converge to largely the same high similarity. Note that the LP and LP-w methods again do quite poorly, except at the high end; the NUL model has no similarity to the known SLC code, as expected.

![Figure 2.5: Adjusted Rand similarity algorithm comparisons](image)
**Jaccard similarity**
The similarity of each algorithm with respect to the SLCs was calculated using the Jaccard similarity, as well as the similarity of a null grouping (NUL) where every student was given the same community. From Figure 2.6, it is clear that of the different approaches and weightings, the SG algorithm has a clearly higher similarity to the known community structure, with the similarity increasing as SLC modularity increases. Interestingly, weighting student links by the number of classes taken together does not help the similarity to actual SLC, indicating that level of student contact is not as important as which students are connected. Again, it can be seen that as modularity increases, all the unassisted approaches converge to largely the same high similarity. Note that the LP and LP-w methods do not perform as poorly as with adjusted Rand similarity, and the NUL model has a higher similarity than most other approaches in the bottom of the SLC modularity distribution. This is because the LP methods and the NUL model each tend to group all students into the same community, and Jaccard similarity does not correct for similarity occurring solely by chance. The conclusion is that when the SLC variable indicates that a high quality community structure exists, the unassisted approaches find it; when the stated communities do not correspond to a high quality community structure, similarity suffers.
**Measuring algorithm speed**

The time an algorithm takes to complete and returns results is a key component of its usability.

There are two important aspects to overall algorithm speed: scalability and base speed.

Scalability measures how many more steps (i.e., how much longer) an algorithm takes for an increase in some aspect of the data size, as the data becomes asymptotically infinite. Base speed has two parts; the first part is a constant amount of setup time to begin the algorithm, and the second part is the multiplicative factor for how long each step takes.

**Measuring scalability**

As the data gets larger, scalability becomes much more important than base speed. The scalability of an algorithm is written with “Big-O” notation and termed running-time. Running
time is how many steps need to be taken to perform the algorithm, ignoring the time for each step. It can be predicted directly by examination of computer code or pseudo code (a simplified sketch of an algorithm). Big-O notation shows how the running time of the algorithm would change given a change in data size. For example, if an algorithm has a running time of $O(n)$, as the data gets very large, it would require one more unit of time for every unit increase in data size, and be called a linear algorithm. Similarly, the fastest typical runtime would be $O(\log n)$, and would be called a logarithmic algorithm; the running time would increase very slowly, with the logarithm of the data size. If the data has two aspects, combining them in notation is also allowed; $O(m \log n)$ denotes an algorithm that increases with the size of data aspect $m$ multiplied by the logarithm of data aspect $n$. Additionally, other combinations of addition and multiplication are allowed. It is important to note that Big-O notation only specifies running time up to a multiplicative factor, which is captured in base speed.

**Measuring base speed**

For relatively smaller data sizes, base speed is most important. If each step takes a long time, an algorithm can take a longer total time even if it has a “faster” Big-O running time than an algorithm with a “slower” Big-O running time, which takes more, but faster, steps.

**Comparing speed**

This analysis measures both aspects of speed. Base speed will be approximated by the average elapsed time to completion. The scalability of each algorithm will be tested empirically, and compared to the scalability predicted by the algorithm’s authors, or by examining the algorithm code or pseudo code.
Comparing scalability
The predicted Big-O running time was found in the approach creator’s papers describing the algorithm. Each algorithm’s scalability was empirically tested using regressions of algorithm elapsed time on a comprehensive set of data aspect size transformations (e.g., \(n\), \(n\) \(m\), \(\log n\), \(n\) \(\log n\), \(m\) \(n\), etc.); the data aspect transformation whose regression explained the most variation in elapsed time was deemed the “measured” scalability, reported, and compared with the predicted value. Note that the empirically measured scalability is sensitive to the particular algorithm implementation used (igraph-0.5.3) (Csárdi & Nepusz, 2006), and the elapsed times are additionally sensitive to the processing speed of the computer (four separate simultaneous single-threaded calculations were run, one thread to a core, on a quad-core processor at 3.6GHz). Finally, several different data aspect combinations may explain similar amounts of variance in elapsed time. For the purpose of the Big-O notation used here, \(n\) will be the number of students in the data, and \(m\) will be the total number of links connecting all students together.

Raghavan, et al. (2007) claims that LP runs in roughly linear time \(O(n)\), although they stop short of providing proof of this. On the range of networks of schools, this algorithm ran much too quickly to allow for accurate estimation of the type of variation in its running time. Pons & Latapy (2005) reported that WT runs in \(O(m n^2)\) in the worst case and \(O(m n \log n)\) typically. On the range of networks of schools, this method runs in \(O(m n)\) time, with 99% of variation explained. Clauset, et al. reported that FG runs in \(O(m n \log n)\) in the worst case and \(O(n \log^2 n)\) typically. On the range of networks of schools, this method runs in \(O(m n \log n)\) time with 95% of variation in time explained. Newman reported that LE runs in \(O((m+n) n^2)\) in the worst case and \(O((m+n) n \log n)\) typically. On the range of networks of schools, this method runs in \(O((m+n) n^2)\) time, with 87% of variation in time explained. Reichardt & Bornholdt did not
specify a running time for SG, and in general, the convergence of the simulated annealing
algorithm depends on the shape of the Hamiltonian energy function for the school network. On
the range of networks in schools, this method runs in $O(m \log n)$ time, with 84% of variance
explained, (95% when removing one large outlier).

| Table 2.4: Predicted vs. implemented scalability of algorithms |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                 | Predicted worst | Predicted typical | As implemented | $R^2$          |
| LP                              | n/a             | $\sim O(n)$     | Very Fast      | n/a             |
| FG                              | $O(m \ n \ \log n)$ | $O(n \ \log^2 n)$ | $O(m \ n \ \log n)$ | 95%             |
| WT                              | $O(m \ n^2)$   | $O(m \ n \ \log n)$ | $O(m \ n)$     | 99%             |
| LE                              | $O((m+n) \ n^2)$ | $O((m+n) \ n \ \log n)$ | $O((m+n) \ n^2)$ | 87%             |
| SG                              | n/a             | n/a             | $O(m \ \log n)$ | 84% (95%)       |

| Table 2.5: Predicted vs. implemented subjective scalability of algorithms |
|---------------------------------|-----------------|-----------------|
|                                 | Predicted scaling | Implemented scaling |
| LP                              | Very good       | Very good       |
| FG                              | Good            | Poor            |
| WT                              | Poor            | Moderate        |
| LE                              | Poor            | Very poor       |
| SG                              | n/a             | Very good       |

Comparing base speed
Turning to a review of base speed, it is very clear that both SG approaches take the longest time.
Figure 2.7 shows that SG times are about 60 times longer than the rest of the approaches. The LP
methods take almost no time, and FG, WT, and LE follow, in order of increasing elapsed time. It
is also clear that the distribution of school size is not uniform across SLC deciles. Calculating
communities using weighted links between students accelerates two methods, (FG, LP), slows
one (SG), and makes no difference to the other (WT), although this is sensitive to
implementation and particular data, of course. Comparing the scaling to the actual speed, even
though the SG algorithms have very good scaling, the slow base speed overcomes the better
scaling for this data size.
**Algorithm efficiency**

To rate the algorithms on efficiency, each quality measure is divided by the elapsed time that it took to achieve that level of quality. Algorithms that achieve decent results in a short period of time are rewarded by this rating. These results are presented in Table 2.6, which is sorted in rough order of increasing efficiency. The LP approaches are the least efficient because of their generally poor performance. The fact that Jaccard similarity will be non-zero even with a random community selection, combined with LP’s extremely fast runtimes, gives LP a misleadingly high efficiency score in that area. Although they have the highest quality, the SG approaches are penultimate in efficiency because of their very slow base speed. FG-w is the most efficient algorithm; the weighted networks accelerate its progress and it produces decent results overall.
Table 2.6: Community detection algorithm efficiency (higher is better)

<table>
<thead>
<tr>
<th></th>
<th>Median Q / t(s)</th>
<th>adj. Rand / t(s)</th>
<th>Jaccard / t(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP</td>
<td>0.000</td>
<td>0.000</td>
<td>13652</td>
</tr>
<tr>
<td>LP-w</td>
<td>0.000</td>
<td>0.000</td>
<td>25111</td>
</tr>
<tr>
<td>SG-w</td>
<td>0.011</td>
<td>0.001</td>
<td>0.006</td>
</tr>
<tr>
<td>SG</td>
<td>0.011</td>
<td>0.003</td>
<td>0.012</td>
</tr>
<tr>
<td>LE</td>
<td>0.322</td>
<td>0.077</td>
<td>0.344</td>
</tr>
<tr>
<td>WT</td>
<td>0.774</td>
<td>0.194</td>
<td>1.391</td>
</tr>
<tr>
<td>WT-w</td>
<td>1.254</td>
<td>0.201</td>
<td>1.301</td>
</tr>
<tr>
<td>FG</td>
<td>2.354</td>
<td>0.406</td>
<td>4.076</td>
</tr>
<tr>
<td>FG-w</td>
<td>5.417</td>
<td>0.767</td>
<td>4.644</td>
</tr>
</tbody>
</table>

Section 2.4: Conclusions

Over all the types of comparisons, one approach stands out for its high quality results: the un-weighted Potts spin glass model (SG). It has the highest quality community structure, Jaccard similarity, and adjusted Rand similarity of all of the algorithms. Its only weakness is its very slow speed and corresponding low efficiency. If time is of no concern, the SG approach is best. The most efficient approach is the fast greedy weighted (FG-w) approach; it performs close to the SG approach on most measures, but runs almost 100 times faster. The WT, WT-w, FG, and FG-w algorithms provide comparable levels of quality; the choice of FG-w is due to its running time: twice as fast as the WT group, and about a third faster than its un-weighted counterpart. The LE approach provides sub-par quality and running times, but the only algorithms to vigorously avoid are the label propagation approaches (LP, LP-w); while fast, on these well-connected, dense school networks they only perform passably on school networks with an extremely high level of community structure.

The SG model has a high rate of accuracy. It classifies more than half of students together and apart when the SLC code does the same, and correctly classifies over three-quarters of students together and apart when the SLC code does the same and SLC community quality is more than 0.3. Further chapters will use the SG model for analysis of community structure.
This chapter has presented a logical definition of community structure, introduced several methods to identify communities of students in schools by using school structure, and benchmarked these algorithmic methods against each other. In short, using unassisted community detection algorithms is feasible for network data about the size of school grades. Higher quality communities measured using the known SLC code are highly correlated with higher quality communities found by the community detection algorithms, and this correlation increases as the community quality of SLCs increases. In addition, the similarity between communities detected by the algorithms increases as SLC code community structure increases, indicating that algorithms are measuring the same thing as SLC coded communities become higher quality. Where a strong SLC structure exists, the algorithms are likely to find the same structure. Where no such SLC structure exists, these approaches still find significant community structure in most schools (modularity > 0.3), indicating that even without SLCs, comprehensive high schools generally have groups of students who take classes together to some extent. While unlike Heck (2004) and Field (2006) this analysis does not attempt to identify classes or organize students into tracks, it also finds groups of students taking classes together, even when using temporal network data instead of multiple year transcripts. The composition of these intra-school groups is addressed in the following chapter.
Chapter 3: Categorizing Community Compositional Variation

This chapter’s aim is to measure and evaluate the composition of communities in schools, and to determine to what degree students in schools and communities are sorted or segregated based on their characteristics (Oakes, 1995; Oakes & Guiton, 1995; D. Ready, et al., 2004). Three primary sorting measures are used to accomplish this aim. The first, assortativity, is a network-theoretic measure evaluating the amount of student-to-student homophily in course-taking patterns (i.e., do similar students tend to take classes together?) (Newman, 2003). The second measure, the Thiel information theory based segregation index (H), measures the amount of diversity across communities in a school (e.g., are all females in one community?) (Thiel & Finezza, 1971). The final measure, the Intra-class Correlation Coefficient (ICC), is used to measure the degree to which academic outcomes of interest vary between communities as compared to within communities (e.g., do all communities have the same average test score?) (Coleman, 1966; McGraw & Wong, 1996).

In this chapter, the student characteristics of interest are described, the three sorting measures are further explained, their results for a range of student characteristics presented, and the results interpreted.

Section 3.1: Description of selected student characteristic variables

The different student characteristics chosen to describe students in this portion of analysis fall into four general categories: demographic characteristics such as gender, ethnicity, and socioeconomic status (SES); course choice characteristics such as taking Advanced Placement (AP) courses; performance measures such as GPA and attendance; and student program participation characteristics such as participation in English as a Second Language (ESL) or the
Gifted program. For the first two sorting measures, assortativity and the segregation index, student categorization is necessary and categories are described below. For the last measure, the ICC, continuous measurements of test scores, GPA, and attendance, are used instead.

**Demographics**
The demographic characteristic variables used in this analysis are gender, free or reduced price lunch status, and ethnicity. Gender and free or reduced price lunch status are binary characteristics. Ethnicity included the following: White, Black, Hispanic, Asian, Filipino, and Other. Ethnicity was included as a categorical variable with all ethnicities, as well as individual binary variables for each ethnicity separate. This is done to determine if particular groups have different experiences than the rest.

**Performance measures**
This category consists of students’ GPA, test scores, disciplinary status, and attendance record. Students are classified by their GPA as follows: students with GPA greater than 3.3 (high GPA) and students with GPA lower than 0.7 (low GPA) are separately compared to everyone else with two appropriately coded binary variables; these particular GPA levels were chosen to represent roughly the top and bottom 10% of annual student GPAs in the population. Their particular value is not essential; the measure’s purpose is to allow the analysis to determine if students who are getting almost all As (or Fails) are taking classes together more so than students in the middle (i.e., are lower performing students structurally together, but higher performing students spread throughout?). Students are also placed into high, medium, and low GPA categories by splitting the students with cut-points at GPA of 1.5 and 2.5 – groups roughly corresponding to about a third of the population each. This measure allows for analysis of the overall distribution of GPA.
Students are split in a similar fashion using the average of their test scores for that year. This measure is different than the test score measure used elsewhere in this research, as students’ test scores are being used here as an overall measure of performance on all tests. Students are classified into the high category if they averaged proficient or above on all tests, and into the low category if they averaged below basic or lower on all state accountability tests taken. The high (proficient or above) and low scoring (below basic or below) students are compared to all other students separately; the three categories are additionally compared with each other (i.e., the distribution of students in the high, medium, and low categories) for a final measure. The distribution of students with disciplinary problems is captured with two variables. The first is whether the student was suspended that year. The second is whether the student was ever suspended in the data available. The final characteristic is the student’s attendance record. Students who are absent more than 20% of the time are assumed to be truant and are compared to all other students.

Program participation
The distribution of students in the gifted, special education, and ESL programs is analyzed as well. These measures help us determine if students in different programs are taking classes together, as expected. Student participation in the gifted and special education programs are separately contrasted to those who are not in the respective program. Students are classified by their ESL status using two methods: the first uses their classification as Limited English Proficient (LEP). The second uses the presence of an ESL class in students’ schedules to identify those students who are still taking specialized classes. Student grouping on these characteristics is expected.
**Student course selections**

This class of student characteristics uses students’ choice of courses to identify to what extent students who select one type of class take all their classes together. The two types of classes analyzed are AP classes and foreign language classes; their corresponding categories include all students who took any AP class or foreign language class, respectively.

**Section 3.2: Assortativity**

Newman (2003) defined assortative mixing in a network to be when nodes “prefer” to be connected to other nodes that are similar to them in some way. Newman hypothesizes that assortative mixing present in a network “will tend to break the network up into separate communities,” giving the example of people preferring to be friends with others that speak the same language, leading to separation of social friendship networks into communities where most people speak one language or the other. This type of separation was also found by a study of a network of telephone calls between speakers of different languages, which found that speakers of the same language formed self-similar communities (Blondel, et al., 2008).

The measure used in this portion of analysis, the assortativity coefficient \( r \), is a network-theoretic measure of the tendency of nodes in a network to have links to other nodes which share the same type (Newman, 2003). The assortativity coefficient is

\[
 r = \frac{\sum_i e_{ii} - \sum \left( \sum_j e_{ij} \right)^2}{1 - \sum_i \left( \sum_j e_{ij} \right)^2},
\]

where \( e_{ij} \) is the fraction of links in a network that connect a node of type \( i \) to one of type \( j \). The assortativity coefficient has a range from negative one to one, where one represents a network showing perfectly assortative mixing, and every node only has connections to nodes that have a particular shared type; zero represents a network showing no assortative mixing at all, every
node is equally likely to connect to a node with a shared characteristic or a node with a different characteristic at random; negative values represent networks where nodes are more likely to be connected to nodes which have different characteristics than could be expected by chance. The assortativity coefficient is undefined if there is only one type of node. This assortativity coefficient is used here on a novel type of network, the network of shared courses between students.

In the high school course taking network, where nodes represent students, and links represent shared classes, assortative mixing is the degree to which students take classes with students who share some specific characteristic. For example, if the chosen characteristic is gender, an assortativity coefficient of zero implies that males and females share classes non-discriminately; an assortativity coefficient of one implies that males only take classes with males, and females only take classes with females. Note that it is very difficult for the assortativity coefficient to reach a high negative value, especially in the high school course-taking network. To see why, note that a coefficient of negative one would imply that every shared class consisted of exactly one female and one male, and have a class size of two, a situation improbable in high school. Similarly, it is difficult for the assortativity coefficient to reach a very high positive value in the high school course taking network for small subgroups; to reach a value of one, there must be enough students of each type to fill at least entire class, and those students must all take classes only with each other – a difficult possibility in a large comprehensive high school of more than two thousand students. This section describes the amount of assortative mixing within the high school course taking networks.
**Results and interpretation**

The assortativity coefficient was calculated for each of the measures over all networks in the dataset, consisting of unique values of school, track, grade, and year. Because the assortativity coefficient is defined on the link level, and schools consist of tens of thousands of links, network-level standard errors calculated using the formula in Newman (2003) similar to that used for Cohen’s kappa yielded network-level standard errors on the order of 10^-5, implying that the assortativity coefficient for any particular network is known very precisely. In addition, the results from this analysis effectively compose a census of the district for these years, and it is possible, not terribly informative to only state, for example, that the assortativity in the district for gender averages exactly 0.02, without a sense of whether individual schools are much higher or lower. To add meaning to the mean, the standard deviation of the coefficients across all schools and networks is calculated. In this descriptive part, qualitatively, this research then assumes that average assortativity coefficients greater than one standard deviation from zero are important (implicitly assuming that the deviation across the entire district represents an estimate of the deviation in the population to which results may be generalized). The number of valid networks (excluding those with no variation in measure, and therefore no valid assortativity), the mean, and the standard deviation of the assortativity coefficient across all networks are presented in each table, grouped by category.

**Course-taking**

All of the measures defined by the types of courses taken by students are more than one standard deviation from zero. In addition, these measures are of higher magnitude than the other measures, with an average coefficient of about 0.16. This result implies that schools in the district are generally keeping students that take Advanced Placement, ESL, or Foreign Language classes together, as also found by Oakes (1985). This is possibly due to scheduling pressures and
the inflexibility associated with taking these specialized classes (Adam Gamoran & Berends, 1987; Hallinan & Sorensen, 1985). For example, schools may have only one section of an Advanced Placement class. Therefore, all students who desire that AP course must take that section. That choice forces a shared class on all those students, creating a shared link and increasing assortative mixing as defined by this measure. As a secondary effect, students that take one Advanced Placement class are more likely to take other advanced classes and identify themselves as students who take advanced classes (Akerlof & Kranton, 2002), a preference shared with their peers who have a shared identity (Hallinan & Smith, 1989; Hallinan & Williams, 1990), also increasing assortative mixing.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class – Advanced Placement</td>
<td>3411</td>
<td>0.23 (0.16)</td>
</tr>
<tr>
<td>Class – Foreign Language</td>
<td>4012</td>
<td>0.13 (0.10)</td>
</tr>
<tr>
<td>Class – ESL</td>
<td>2709</td>
<td>0.18 (0.15)</td>
</tr>
</tbody>
</table>

**Special programs**

Assortativity is also higher among students in special programs, such as special education or English as a Second Language. The high school course taking network shows assortative mixing among students that are taking ESL classes, and among those designated as Limited English Proficient to a lesser extent. Students in the special education and gifted programs also tend to share classes, although the gifted program shows less mixing of students. These results are broadly consistent with other research that indicates that tracking in certain forms is still active in schools (Adam Gamoran & Berends, 1987; Oakes & Guiton, 1995).
Table 3.2: Average school assortativity by student program participation

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gifted</td>
<td>4040</td>
<td>0.09 (0.10)</td>
</tr>
<tr>
<td>Special Education</td>
<td>3953</td>
<td>0.12 (0.08)</td>
</tr>
<tr>
<td>English Learner – LEP</td>
<td>3976</td>
<td>0.14 (0.11)</td>
</tr>
<tr>
<td>Class – ESL</td>
<td>2709</td>
<td>0.18 (0.15)</td>
</tr>
</tbody>
</table>

**Academic outcomes**
The high school course taking networks show some evidence of assortative mixing on demonstrated ability. Students with similar test scores and annual GPAs tend to take classes together, along with students together in the gifted and special education programs, an unsurprising result. Students that receive very good marks are somewhat more likely to be taking classes with students that also receive very good marks, but the converse is not well supported by this evidence. While the absolute magnitude of the average assortativity is small, this analysis provides quantitative support for the existence of groups of students with very different academic outcomes within schools that are not formally tracked – SLCs were supposed to have an “equitable” distribution of students (Price, et al., 2003; Rayyes & Barela, 2009). This result is consistent with the research consensus that ability-based “tracks” exist and can form without clear administrative direction due to the system of prerequisites and scheduling pressures in the modern school (Adam Gamoran & Berends, 1987; Heck, et al., 2004; Neubig, 2006; Oakes & Guiton, 1995; Powell, et al., 1985).
Table 3.3: Average school assortativity by student academic outcomes

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Scores</td>
<td>3547</td>
<td>0.07 (0.06)</td>
</tr>
<tr>
<td>High Average Test Score</td>
<td>3545</td>
<td>0.07 (0.09)</td>
</tr>
<tr>
<td>Low Average Test Score</td>
<td>3547</td>
<td>0.09 (0.08)</td>
</tr>
<tr>
<td>GPA</td>
<td>4072</td>
<td>0.06 (0.04)</td>
</tr>
<tr>
<td>High Annual GPA</td>
<td>4072</td>
<td>0.06 (0.05)</td>
</tr>
<tr>
<td>Low Annual GPA</td>
<td>3818</td>
<td>0.03 (0.03)</td>
</tr>
</tbody>
</table>

**Attendance and disciplinary actions**

There is no significant evidence of any differences in class association between students that are truant or students that are subject to disciplinary actions. These students are spread across the school, and do not take classes together to a degree much greater than chance. It is important to note that inference from this result is limited because assortativity is subject to attenuation when only a small proportion of students have the studied characteristic, as is true in this case – few students are suspended or miss large amounts of school.

Table 3.4: Average school assortativity by student attendance and discipline

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truant</td>
<td>3939</td>
<td>0.02 (0.02)</td>
</tr>
<tr>
<td>Discipline – Annual</td>
<td>3979</td>
<td>0.01 (0.02)</td>
</tr>
<tr>
<td>Discipline – Overall</td>
<td>4050</td>
<td>0.02 (0.02)</td>
</tr>
</tbody>
</table>

**Demographics**

There is no significant evidence of any within-school assortative mixing on ethnicity, gender, home language, or SES, as measured by free and reduced price lunch status in high school. While inference is limited, this is consistent with the hypothesis that, on average, more sorting inside high schools is on the basis of other factors and that sorting on demographics is too diffuse to effectively measure with assortativity. This is partly because most schools in Los Angeles have very high percentages of Hispanic students, attenuating results for other races, but it is also an example of a change since the 1970s, when sorting on race occurred in schools to a greater extent (Farley & Taeuber, 1974). Now, these results share more in common with those by Clotfelter (1999) who found that most modern segregation was between districts or between
schools, and not within districts or schools within districts (Clotfelter, 1999), perhaps as a result of years of residential sorting (Denton, 1995; Lankford & Wyckoff, 2006; Tiebout, 1956).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>4072</td>
<td>0.02 (0.03)</td>
</tr>
<tr>
<td>Free or Reduced Price Lunch</td>
<td>4029</td>
<td>0.01 (0.03)</td>
</tr>
<tr>
<td>All Ethnicities</td>
<td>4029</td>
<td>0.04 (0.04)</td>
</tr>
<tr>
<td>Ethnicity – Black</td>
<td>3763</td>
<td>0.02 (0.03)</td>
</tr>
<tr>
<td>Ethnicity – White</td>
<td>3220</td>
<td>0.02 (0.04)</td>
</tr>
<tr>
<td>Ethnicity – Hispanic</td>
<td>4029</td>
<td>0.05 (0.06)</td>
</tr>
<tr>
<td>Ethnicity – Asian</td>
<td>3221</td>
<td>0.03 (0.05)</td>
</tr>
<tr>
<td>Ethnicity – Filipino</td>
<td>2938</td>
<td>0.01 (0.02)</td>
</tr>
</tbody>
</table>

**Limitations**

The primary limitation of this presentation of results is that viewing the average assortativity coefficient across the district obscures significant sorting differences between schools, due to Tiebout sorting (Denton, 1995; Lankford & Wyckoff, 2006; Tiebout, 1956). The standard deviation of the assortativity coefficient indicates that some schools have much higher levels of assortativity than the average; as mentioned before, given the high number of links, the assortativity is calculated very precisely at the network level. Preliminary analysis on individual schools indicates that many schools show highly statistically significant assortativity qualities individually. Future analysis should investigate whether the school networks with high and low assortativity have different average outcomes than other schools, or have different scheduling or course offerings.

Another limitation associated with this presentation of the assortativity coefficient is that which is necessarily associated with calculating and interpreting the average of any quantity with missing values. Since the assortativity is undefined with only one community or one type of student, schools that have either of these qualities are excluded from the average, as can be seen
in the number of schools. Additionally, and potentially exacerbating this weakness, school networks with even just one student with a particular characteristic are given equal weight as networks with many students of all types. This analysis gives each network an equal weight. Future analysis should explore how other weighting schemes, such as weighting by the number of students, change results.

A final limitation is that, as noted in an earlier example, these networks are formed from shared classes. Therefore, the maximum assortativity coefficient possible may not be one; it depends on the percentage of the population with a selected characteristic or other school measures. For example, suppose there are five students with a characteristic; if the smallest class size is fifteen students, those students must take classes (and therefore form links) with at least ten other students who have different characteristics, shrinking the magnitude of the maximum possible assortativity towards zero. Because this is a novel use of the assortativity coefficient to provide information about school sorting and structure, no prior research exists on use of this measure in a similar context exists, and this analysis makes no attempt to derive, quantify, or correct for this likely downward bias, saving those efforts for further research.

**Conclusion**
The novel use of a network-theoretic measure provided insight on some grouping patterns within schools. Some results were easily anticipated (e.g., special education students take classes together); others, less so (e.g., apparent lack of sorting on gender or ethnicity). In general, though SLCs were supposed to have equitable admissions, while the amount is slight, there is some evidence for grouping or tracking by ability and interest, as was also found by many researchers (Adam Gamoran & Berends, 1987; Heck, et al., 2004; Neubig, 2006; Oakes, 1995; Powell, et al.,
There was no evidence that attendance, discipline, or ethnicity have any effect on grouping, though inference from this result should be tempered by the potential downward bias (attenuation) in these measures due to the relatively small numbers of students with certain ethnicities, attendance or discipline problems in certain schools. It is also important to note that this analysis does not attempt to determine the changes in these measures after SLC implementation, and so no inference about SLC effects or correlations can be drawn.

Section 3.3: Segregation

The implementation of SLCs may change the amount of heterogeneity in the composition of communities of students who take classes together. To measure this, a tool from the discipline of sociology is used. Researchers have used different segregation measures to define and quantify inequality among the distribution of groups in an organization (James & Taeuber, 1985; Massey & Denton, 1988; White, 1986). While such measures of heterogeneity have been applied to schools and districts (Freeman, 1978; Thiel & Finezza, 1971; Zoloth, 1976), as well as cities and census blocks (Bayer, McMillan, & Rueben, 2004; Coleman, 1966; Duncan & Duncan, 1955; Farley & Taeuber, 1974; White, 1983), this analysis is a novel application of these measuring tools of sociology to the communities formed by students taking classes together within schools.

The Thiel information theory segregation Index

This analysis uses a measure developed by Henry Thiel to measure the amount of heterogeneity in the composition of groups within schools (Thiel & Finezza, 1971). This index is used because a comprehensive review, (Reardon & Firebaugh, 2002), of a set of more than six different multi-group segregation indices (James & Taeuber, 1985; Massey & Denton, 1988; White, 1986; Zoloth, 1976), it was the only index that obeyed the principle of transfers, always decreasing when an individual of type j is moved from a group with a high proportion of similar individuals
to a group with a low proportion of individuals of type j. It was one of the two segregation measures which allow for its decomposition into pieces whose contribution is also defined and measurable. The Thiel information theory index (H) has a range from zero to one, inclusive: a value of one represents full segregation (i.e., homogeneous subunits) and a value of zero represents no segregation (evenly heterogeneous subunits).

**Results and interpretation**
The question of interest is: what is the level of heterogeneity within schools across groups of students taking classes together? A high level of this segregation index indicates that communities found within schools have different average levels of the characteristic on which the index is defined. A low level of this index indicates that the composition of community groups within schools is representative of the composition of school population as a whole. This analysis is not concerned with segregation or the level of differences across schools in the district, only the level of heterogeneity of communities of students within schools. Therefore, the district level average of these measures for each of the student characteristics is calculated for student communities found algorithmically, and the mean and standard deviation are reported and interpreted by group. This analysis yields results broadly similar to those for assortativity.

**Course-taking**
The strongest result for segregation within schools is along students’ choices of classes taken. Students that take AP classes, ESL classes, and foreign language classes tend to be more concentrated in communities. This is partly influenced by the fact that communities were created by analyzing course-taking patterns, and these types of classes are those that are relatively rare with few sections offered, and are taken by students who are likely to share classes for two other reasons: inflexibility of the master schedule, and student desires (Heck, et al., 2004; Neubig,
This might explain the evidence that students are sorted into “AP student” and “foreign language” communities.

### Table 3.6: Average school community segregation by student course choices

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class – Advanced Placement</td>
<td>4060</td>
<td>0.24 (0.20)</td>
</tr>
<tr>
<td>Class – Foreign Language</td>
<td>4060</td>
<td>0.11 (0.10)</td>
</tr>
<tr>
<td>Class – ESL</td>
<td>4060</td>
<td>0.19 (0.20)</td>
</tr>
</tbody>
</table>

**Program participation**
Communities of students tend to be more homogeneous with regard to student program participation. Within schools, different communities tend to have different levels of gifted students, special education students, and ESL students, a result also found by analyzing pair-wise shared classes in the assortativity analysis. This implies that, in general, within schools, groups of students who share classes tend to also share program participation status, instead of students in these programs being evenly distributed across the groups within the population. This result is again similar to that found in research. Namely, students sort themselves into groups who share characteristics (Adam Gamoran & Berends, 1987; Heck, et al., 2004; Oakes & Guiton, 1995).

### Table 3.7: Average school community segregation by student program participation

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gifted</td>
<td>4060</td>
<td>0.16 (0.11)</td>
</tr>
<tr>
<td>Special Education</td>
<td>4060</td>
<td>0.22 (0.17)</td>
</tr>
<tr>
<td>English Learner – LEP</td>
<td>4060</td>
<td>0.18 (0.15)</td>
</tr>
<tr>
<td>Class – ESL</td>
<td>4060</td>
<td>0.19 (0.20)</td>
</tr>
</tbody>
</table>

**Academic outcomes**
As seen with assortativity, the high school course taking network also shows a tendency to differentiate between students of different measured ability. This analysis shows that communities of students are at least somewhat separated based on average level of test scores and student grades earned, with the coefficient for test scores being greater than that of grade point average. This implies that students who tend to take classes together also tend to have similar levels of measured ability, which was also found by Heck (2004 and Field (2006), among
others. Interestingly, the coefficient for GPA is less than that for test scores, consistent with the hypothesis that teachers grade at least partially on a curve (McMillan, 2001).

**Table 3.8: Average school community segregation by student academic outcomes**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Scores</td>
<td>3093</td>
<td>0.13 (0.08)</td>
</tr>
<tr>
<td>High Average Test Score</td>
<td>3093</td>
<td>0.17 (0.11)</td>
</tr>
<tr>
<td>Low Average Test Score</td>
<td>3093</td>
<td>0.14 (0.09)</td>
</tr>
<tr>
<td>GPA</td>
<td>4060</td>
<td>0.09 (0.05)</td>
</tr>
<tr>
<td>High Annual GPA</td>
<td>4060</td>
<td>0.10 (0.06)</td>
</tr>
<tr>
<td>Low Annual GPA</td>
<td>4060</td>
<td>0.10 (0.08)</td>
</tr>
</tbody>
</table>

**Discipline and attendance**

Unlike the results from analysis of assortativity, there are differences across communities in the average proportion of students who have had disciplinary problems, as well as the proportion of students who are designated as habitually truant. So, while as a whole, students with these risky behaviors tend to take classes evenly across students (as shown by assortativity), and the implementation of SLCs might reduce these behaviors in the future (Klonsky, 2002), there are communities of students that never display these behaviors, and communities that tend to have relatively higher levels of students who have risky behaviors, a result that Ready (2004) also noted. The average coefficient of 0.08, compared to the potential maximum value of one, still implies a fairly even distribution across communities, however.

**Table 3.9: Average school community segregation by student attendance and discipline**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truant</td>
<td>4060</td>
<td>0.09 (0.06)</td>
</tr>
<tr>
<td>Discipline – Annual</td>
<td>4060</td>
<td>0.08 (0.06)</td>
</tr>
<tr>
<td>Discipline – Overall</td>
<td>4060</td>
<td>0.06 (0.05)</td>
</tr>
</tbody>
</table>

**Demographics**

While some demographic segregation within schools is present, it is of lesser magnitude than the segregation which exists with some other student characteristics. Within school course-taking community segregation by gender or socioeconomic status – as measured by free or reduced price lunch eligibility – is effectively non-existent, contrary to what is predicted by older
research (Erbe, 1975). Segregation along home language is even smaller in magnitude and not shown here. While ethnic segregation is the largest of this set of measures and the Asian group is most segregated, at 0.11 their segregation is still smaller than test score segregation, and comparable to segregation on GPA. Still, having some segregation is consistent with the combination of student choice and homophily, which can help create this type of self sorting inside schools (Kubitschek & Hallinan, 1998; McPherson, Smith-Lovin, & Cook, 2001; Moody, 2001), even as the level of segregation overall decreases (Clotfelter, 1999).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>4060</td>
<td>0.03 (0.04)</td>
</tr>
<tr>
<td>Free or Reduced Price Lunch</td>
<td>4060</td>
<td>0.03 (0.03)</td>
</tr>
<tr>
<td>All Ethnicities</td>
<td>4060</td>
<td>0.11 (0.07)</td>
</tr>
<tr>
<td>Ethnicity – Black</td>
<td>4060</td>
<td>0.09 (0.09)</td>
</tr>
<tr>
<td>Ethnicity – White</td>
<td>4060</td>
<td>0.09 (0.10)</td>
</tr>
<tr>
<td>Ethnicity – Hispanic</td>
<td>4060</td>
<td>0.08 (0.07)</td>
</tr>
<tr>
<td>Ethnicity – Asian</td>
<td>4060</td>
<td>0.11 (0.10)</td>
</tr>
<tr>
<td>Ethnicity – Filipino</td>
<td>4060</td>
<td>0.08 (0.09)</td>
</tr>
</tbody>
</table>

**Limitations**

The primary limitation of this analysis is the use of the district-wide average level of heterogeneity. Using the average value of an index can obscure individual schools with much greater or smaller levels of the index. At certain schools and grades, more intense segregation certainly exists, as shown by the size of the standard deviation of segregation measures. This limitation is somewhat ameliorated by the district level focus of this research.

Another limitation associated with this use of the average is that all school-grade-year combinations are given equal weight, regardless of the number of students of any particular type present (i.e., this average is not weighted by the number of students). This might lead to a downward bias in the average level of segregation, as school networks that are homogeneous
with regard to a particular characteristic are merged into the average and treated as having zero segregation. Finally, these indices are calculated based on communities determined by a community detection algorithm. The indices are only as valid as the results of the algorithm.

**Conclusion**

Generally, the pattern of results for this segregation measure matches that of assortativity, but with somewhat stronger results (as measured by effect size). This might be because assortativity is a simple simply pair-wise measure, and these results are based on an algorithm that uses all classes taken by students, with the flexibility to ignore classes that do not contribute to the community formation. Still, on average, there is little absolute segregation present in the high school course taking network within schools in this school district. However, segregation across schools, while not the focus of this research, can be presumed to exist due simply to residential sorting (Denton; Lankford & Wyckoff, 2006; Tiebout, 1956).

Relatively more segregation within schools is based on students’ choices of classes taken, students’ participation in special programs, and students’ demonstrated ability, a result that is broadly consistent with research indicating that, left to their own devices, for many possible reasons, students in schools display *homophily*, or self-sorting, into tracks (Adam Gamoran & Berends, 1987; Heck, et al., 2004; Neubig, 2006; Oakes & Guiton, 1995; Powell, et al., 1985; D. Ready, et al., 2004). To the degree that this is desired sorting, students and teachers might be getting a net benefit (Bryk & Driscoll, 1988). While the results indicate that certain schools have high levels of within-school segregation, students in general attend schools with integrated environments across all sorting characteristics – possibly due, in some part, to aforementioned attenuation issues. It is important to note that even the highest average segregation characteristic, that of students taking Advanced Placement classes, is only 0.24, still a small amount compared
to the index’s potential maximum of one. In this district, the distribution of student types in
groups of students who take classes together (structural student communities) is largely similar
to that of the school to which that group belongs.

**Section 3.4: Intraclass Correlation Coefficient**

Both previous sorting measures used cutoff points to analyze the underlying continuous variables
of GPA, test scores, and attendance. Evaluation using those cutoffs might produce results that
are sensitive to the particular cutoffs used. As a sensitivity check, a different tool, the intraclass
correlation coefficient (ICC), is used to analyze the variation of these characteristics across
communities directly. The ICC is an estimate of the proportion of variation in the chosen
outcome that is explained by a grouping at a selected level (McGraw & Wong, 1996).

There are many different types of ICCs (McGraw & Wong, 1996). For the purposes of this
analysis, however, we use the ICC derived from an unconditional hierarchical linear model to
describe the proportion of variance explained at the community level of analysis for the selected
continuous outcomes (Bryk & Raudenbush, 1988; Raudenbush & Bryk, 2002).

**The hierarchical model**

For the purpose of this analysis, the value of interest is the proportion of variance in outcomes
explained by communities of students. Therefore, a hierarchical model of levels above the
community level and corresponding decomposition of the outcome variance into additional
school-level, grade-level, or year-level portions is unnecessary. However, the effect of cross-
school variance in outcomes must be removed, or it would inflate the within-school estimates of
variation explained. To do this, a two-level model is formed. The first level is all the
combinations of groups formed above the community level (i.e., all combinations of school,
The base of the hierarchical null random effects model is as shown in Equation 3.1. The student level outcome, $Y$, is assumed to be composed of random individual residual variation ($I$) around a community level mean.

**Equation 3.1: Student level variance**

$$Y_{ics} = \beta_{0ics} + I_{ics}$$

$I \sim N(0, \sigma_i^2)$

As shown in Equation 3.2, the community level mean is composed of random residual variation ($C$) around a grade-track-year-school level mean.

**Equation 3.2: Community level variance**

$$\beta_{0ics} = \beta_{00s} + C_{cs}$$

$C \sim N(0, \sigma_c^2)$

Finally the grade-track-year-school level mean is composed of random residual variation ($S$) around the grand mean. Each of the residuals, $I$, $C$, and $S$, is assumed to be normally distributed with mean zero and individual variance to be estimated by a restricted maximum likelihood procedure.

**Equation 3.3: Upper level variance**

$$\beta_{00s} = \mu_{000} + S_s$$

$S \sim N(0, \sigma_s^2)$
These three levels form the null hierarchical model, with students nested in communities which are nested in the combination of school, grade, and year. Because this analysis focuses on the community level, the intra-class correlation of interest is the fraction of the total variance of I, C, and S, which is explained by the community membership of a student (C), as shown in Equation 3.4.

\[
\text{Equation 3.4: The community level ICC} \quad \frac{\sigma_c^2}{\sigma_i^2 + \sigma_c^2 + \sigma_s^2}
\]

This community level ICC can be interpreted as the amount of difference in the outcome across communities, as compared to within communities or across schools. An ICC of 1.0 would imply that 100% of the variation in the outcome is explained by differences in community level average outcomes and, correspondingly, that there is no difference across schools or students. An ICC of 0.0 would imply that the community groupings are meaningless, and that 0% of the difference in outcomes can be explained by the community level of grouping; therefore, variation in outcomes is either at the individual level or the school level, or some combination of the two. ICCs between these two extremes would imply that the community membership of a student explains some but not all of the variation in student performance.

**Student characteristics**
This analysis uses three measures separately, the standardized English test score (z-score), the GPA, and the percentage of days attended, as its continuous measures.

**Results**
The ICC at the community level was calculated for the standardized English test scores, annual GPA, and attendance. Results are presented in Table 3.11.
Table 3.11: Variance decomposition of test scores, grades, and attendance

<table>
<thead>
<tr>
<th>Student characteristic</th>
<th>Student level variance ($\sigma_i^2$)</th>
<th>Upper level variance ($\sigma_s^2$)</th>
<th>Community level variance ($\sigma_c^2$)</th>
<th>Community ICC %</th>
</tr>
</thead>
<tbody>
<tr>
<td>English test score</td>
<td>0.66</td>
<td>0.11</td>
<td>0.27</td>
<td>26%</td>
</tr>
<tr>
<td>GPA</td>
<td>0.76</td>
<td>0.088</td>
<td>0.14</td>
<td>14%</td>
</tr>
<tr>
<td>Attendance</td>
<td>0.0085</td>
<td>0.00077</td>
<td>0.00072</td>
<td>7%</td>
</tr>
</tbody>
</table>

**Interpretation**

**Test scores**
As Table 3.11 shows, the community level is accountable for 26% of variation in students’ test scores, with the vast majority of the remaining variance at the student level, and not the upper level, which includes the across-school level, a result also found in the Coleman report (Coleman, 1966). Compared to the upper level variance, the community level within schools explains more than twice as much student performance as all differences across all upper levels. This means that, on average, switching from the highest performing community in a school to the lowest performing community in a school can result in a change comparable to more than twice the magnitude of switching from the highest performing combination of school, track, grade, and year to the lowest combination of these. This implies that there are generally higher and lower performing groups within schools, as posited by Ready (2004), and that for all the difference that exists across schools, as in the Coleman report and Hanushek (1998), this analysis finds that there is more variation within schools and students (Coleman, 1966; Hanushek, Kain, & Rivkin, 1998).

**Annual student GPA**
The community level is accountable for 14% of the variation in students’ grades, explaining variation in students about as well as all upper level variance. Even more of the variation in GPA is explained at the student level, as can be expected; schools tend to give a similar distribution of grades to their students regardless of level of test scores. However, we see that, again, the
amount of variation within communities in schools is more than that variation across all upper levels. Given the varying levels of mean test score performance across schools, this suggests that certain schools should have a higher expected GPA given students’ test scores (i.e., there is evidence that teachers grade on a curve) (McMillan, 2001; Starch & Elliott, 1912), although this hypothesis is not tested here.

**Attendance**
The community level is accountable for about 7% of the variation in students’ attendance. While this is about equal to the variation explained by the upper levels, student level variation is much larger. Attendance appears to be a measure that is student-driven and not related to types of course taking; communities all have fairly similar levels of student attendance. There is also not much variation in attendance across schools.

**Limitations**
The primary limitation of this method is that the hierarchy of levels above the community levels has been collapsed to independent units. While this leaves the estimate of community level variation intact and sidesteps the issue of estimating crossed or mixed effects at several levels, differences within or between higher levels cannot be estimated. This somewhat reduces the interpretability of results (e.g., this analysis cannot determine the amount of variation at the school level or year level separately). The decision to combine all upper levels was made for estimation purposes, as maximum likelihood (ML) methods tend to fail when estimating a true variance of zero at any level, and some of the levels had an estimate of variance that was nearly nil, confirmed by using an expectation-maximization (EM) algorithm in place of the ML algorithm.
Another weakness of this method is the assumption of normality on the dependent variable and on the residuals (random effects or error terms) at each level. Unfortunately, the assumption of normal residuals cannot be tested, and is left in place as a caveat to these results. The assumption of normality of the dependent outcome can be tested, and was satisfied in the case of English test z-scores (by construction) and annual GPA. It is not satisfied, however, in the case of percent of school days attended, as that variable is strongly skewed towards the right (100% attendance), so those results, while presented, should be interpreted with caution. Finally, this analysis relies on the validity of community identifiers, whether determined by a community detection algorithm or entered by a counselor by hand.

Conclusion
The decomposition of variance analysis showed that differences across communities in a given grade level are important – roughly similar to those differences that exist across all upper levels combined. However, the student level variation is a much larger factor in all of the three measures: test scores, attendance, and GPA. This fact, combined with the ratio of community-level to upper-level variation, suggests that changes to a student’s group of peers inside a school can be as significant a change in environment as moving from the highest performing school to the lowest performing school in the school district. These results are largely compatible with Hanushek (1998), the Coleman report (1966), and the research consensus that regularly finds that the individual student level is the most important level of variation, and with within-school factors explaining large amounts of residual variance.

Section 3.5: Summary
This chapter has described the type of variation in student characteristics that exists in the high school student level course-taking network across the district. The analyses at the student and
community levels have similar, complementary results. The district has little variation or segregation within school communities on average. However, this conclusion comes with two exceptions. First, there are certainly schools that show very large amounts of segregation and variation between communities, as predicted by Tiebout (1956) and Denton (1995). Second, analysis of ability and program characteristics indicates that there is some modest grouping based on ability and special program participation. The research consensus is still unclear as to a causal direction of effects of this type of ability segregation on students (Carbonaro, 2005; Chiu et al., 2008; Martin, Karabel, & Jaquez, 2005).

In short, while most students attend classes in an environment that resembles the school as a whole, students in the ESL, Gifted, or Special Education programs tend to take classes together, as shown by Heck (2004) and Field (2006). To a lesser degree, student course-taking communities show grouping by ability: students with higher or lower GPA or test scores tend to be in the same course-taking group as other similar students, as predicted by Ready (2004). Finally, the differences between communities inside schools are larger than the differences between schools, regardless of the chosen outcome measure, as predicted by Hanushek (1998) and Coleman (1966), among others.
Chapter 4: School Structure Changes across Time

Section 4.1: What is the expected effect of SLC implementation on school structure?

The implementation of the whole-school SLC policy should create changes in the courses that students take. Due to the intentional grouping of teachers and their respective classes into communities (Price, et al., 2003), the effective size of the community of students that a student sees every day should decrease (Cotton, 2001). This decrease in effective community size is associated with an increase in the feelings of connectedness for those students (McNeely, Nonnemaker, & Blum, 2002; Weiss, Carolan, & Baker-Smith, 2010). Additionally, the student may share course schedules with more students (Neubig, 2006), and see each other more often during the day, giving them more of an opportunity to make friends (Bryk & Driscoll, 1988), increase their attachment to school (Shear, et al., 2008), and thereby improve academic outcomes (V. E. Lee & Smith, 1993).

This chapter focuses on the effect of the implementation of whole-school small learning communities on quantitative structural measures associated with the environment that students face within the school. The specific structural elements analyzed in this analysis are the average number of students in students’ class neighborhoods (degree), the average number of classes taken with each student, and the overall level of community structure present in the community (modularity).

Example SLC implementation scenario
As a descriptive example of types of school structure, let us examine two students’ course-taking experiences in a hypothetical large comprehensive high school. Each student takes six hours, or
periods, of class each day, and each class has exactly twenty-one students. Examining each student, Student A has a more traditional comprehensive high school experience, where each class consists of a new set of students each period, and no student is seen twice in class throughout the day, for a total of 120 different students each in his class neighborhood. Student B has a “cohort” experience, and sees the same set of twenty other students each period, for a much smaller class neighborhood.

<table>
<thead>
<tr>
<th>Table 4.1: Sample extreme scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of students seen during the day (class neighborhood)</strong></td>
</tr>
<tr>
<td>Average number of classes taken with each other student</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Student A has little incentive to make friends with students in class, because he does not see any student in class more than once each day. Instead, Student A would be better off making connections with other children around activities that do not involve school, or involving lunch or other non-scholastic meeting times in school. If Student A misses classes, and would like to catch up on homework by calling friends, he would need to call six different people to get information on all classes he is taking.

On the contrary, since Student B shares all six classes with other students, there are many opportunities for interaction and friend formation in and surrounding class. In addition, the payoff to becoming friends with other students in class is high; that student would see their friends all day. If Student B is absent, she could call one person (who is also possibly a friend) and receive information on what happened in all classes. Group work and homework study sessions would also be easier to plan. Any classroom influences would be magnified throughout the day, as the same students meet hour after hour. Hopefully, teachers would be able to see
these influences developing and work with the other teachers to support the positive influences and suppress the negative influences.

A similar analogy can be drawn for teacher relationships, as suggested by economic and business organization literature (Lesser & Storck, 2001; Reagans & McEvily, 2003; Wenger, 1998). Assuming that teachers have a limited amount of time to make professional connections at school over years and during the day, teachers that regularly share a cohort of students would receive greater utility from, and have higher returns to, their efforts to collaborate on ways to deal with problem students or support and boost high-achieving students, than teachers that do not share many students. This type of collaboration can occur over the long term, for example, where teachers use a portion of their assigned professional development days to talk about their students and, for example, possibly discover that a particular student has an uncorrected vision problem, and move them to the front of the classroom. The collaboration facilitated by sharing a smaller cohort can also involve more transient issues, such as a teacher sharing information with their colleagues over the passing period concerning issues to watch out for; for example, suppose that two students recently had an intense argument, the teacher can be vigilant when the students arrive and be prepared to prevent the situation from escalating into a fight (Febey, 2006). Or, suppose a student came to class emotionally distraught; that information could be passed along and each teacher could watch and provide support through the day, increasing students’ feelings of connectedness (McNeely, et al., 2002).

In practice, SLCs would lie between these two extremes. It is not the stated goal of SLCs, and indeed, it is impossible, given a typical school master schedule (Johns, 2008b), for schools to
create cohorts of 20 or 30 students that stay together through the entire day. Instead, SLCs can realistically aim to provide somewhat larger groups of students that share some classes, but smaller than groups that would otherwise exist (Neubig, 2006). SLCs also generally try to allow students to take “passport” classes, classes outside of the core group of classes (D. Ready, et al., 2004). These “passport” classes are usually advanced classes or electives.

Section 4.2: Definitions of school structural measures

Average size of a student’s community
The student class neighborhood is defined to be all other students one student shares classes with. When a school grade implements wall-to-wall SLCs, this number of students should decrease. In the network formulation, this corresponds to the degree of the student (Freeman, 1977). This is a student-level measure.

Average number of classes taken with each student
This measure is the average number of classes taken with each other student in a student’s class neighborhood. When a school implements SLCs, this number should increase. In the network formulation, this corresponds to the weighted degree divided by un-weighted degree (Freeman, 1977). This is a student-level measure.

Maximum community structure found
This measure is from Chapter 2. A set of five community detection algorithms were run on the dataset of courses taken by every student, and the maximum modularity (Newman & Girvan, 2004), or amount of community structure, found by the highest quality algorithm is reported, for each grade, school, semester, and year. The implementation of SLCs should increase the level of community structure found in the student affiliation network. This is a school-level measure.
Section 4.3: Modeling approach

The model used for the analysis is a comparative interrupted time series model (Shadish, Cook, & Campbell, 2002). The variation in implementation date both within and across schools is used to estimate the effect of implementing SLCs across an entire grade (Bloom, 2003; Bloom & Riccio, 2005). Because schools sometimes implemented SLCs in some grades before others, each grade is considered separately. Each school and grade is used as its own control. Schools that did not experience a change in SLC implementation status are effectively excluded from the model. Community structure measures are averaged at the student-year level for student-level measures, and the school-year level for school level measures (average measures across tracks and semesters are used). The base specification used to model the effects of SLCs is Equation 4.1. The coefficient on a grade by year by school matrix of indicators of SLC implementation is the causal effect of SLC implementation on that measure. Errors are clustered by school and year to account for any year-based school-level error factors, such as data entry errors, and the fact that students are generally nested within schools.

Equation 4.1: Model specification 1: Base structural model

\[ Y_{sgt} = \beta (\text{SLC})_{sg} + \alpha_{sg} + \mu_{ist} \]

\[ \mu_{ist} = \epsilon_{st} + \epsilon_{i} \]

*The meaning of symbols for all of these models is as follows:*

- \( Y \) = student or school level structural measures
- \( X \) = vector of school sizes
- \( s \) = school
- \( g \) = grade
- \( t \) = time
- \( i \) = students

*School-grade fixed effects*

*Errors clustered by school-year*

Additional specifications

To capture any district-wide year-based variation in course requirements or structure, a set of year indicator variables is added to control for any district-wide individual year-based variation (Equation 4.2). Because year is correlated with SLC implementation date, however, the year may
absorb some of the effect of any partial SLCs created before a school grade completes wall-to-wall implementation of SLCs, causing shrinkage towards zero.

**Equation 4.2: Model specification 2: Year controls**

\[ Y_{ist} = \delta (SLC)_{sgt} + \alpha_i + \alpha_{sg} + \mu_{ist} \]

\[ \mu_{ist} = \varepsilon_{st} + \varepsilon_i \]

Because school size is also related to these structural measures, an additional specification with school size as a control is added (Equation 4.3).

**Equation 4.3: Model specification 3: Size controls**

\[ Y_{ist} = \delta (SLC)_{sgt} + \beta X_{ist} + \alpha_{sg} + \mu_{ist} \]

\[ \mu_{ist} = \varepsilon_{st} + \varepsilon_i \]

Finally, a combination of these specifications is used for the full model (Equation 4.4).

**Equation 4.4: Model specification 4: Year and size controls**

\[ Y_{ist} = \delta (SLC)_{sgt} + \beta X_{ist} + \alpha_i + \alpha_{sg} + \mu_{ist} \]

\[ \mu_{ist} = \varepsilon_{st} + \varepsilon_i \]

**Section 4.4: Distributional descriptive statistics**

The initial summary distributional characteristics of these measures in the 2003-2004 school year are presented in Table 4.2. From analysis of finely grained histograms and these measures, the distribution of number of classes taken with each student seen is bounded by one and the mean is skewed towards the right of the distribution; the other two measures are not skewed much visually, and appear roughly normal.

<table>
<thead>
<tr>
<th>Table 4.2: Structural descriptive statistics in 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Number of other students</td>
</tr>
<tr>
<td>Classes taken with each student</td>
</tr>
<tr>
<td>Community Structure Found</td>
</tr>
</tbody>
</table>

The distributional characteristics of these measures in the 2007-2008 school year, after many schools implemented SLCs, are in Table 4.3. Note again the number of classes taken with each student seen is again skewed higher. All measures changed, with community structure and number of classes taken with each student increasing, and number of students seen each day decreasing.
Table 4.3: Structural descriptive statistics in 2008

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of other students</td>
<td>126.6 (34.1)</td>
<td>127.5</td>
</tr>
<tr>
<td>Classes taken with each student</td>
<td>1.46 (0.53)</td>
<td>1.27</td>
</tr>
<tr>
<td>Community Structure Found</td>
<td>0.39 (0.087)</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Section 4.5: Model results

The results from the econometric model in Table 4.4 indicate that the implementation of SLCs did create a change in these elements of school structure. Specification 1 tests the base model with school and grade fixed effects. Specification 2 adds year controls. Specification 3 replaces the year controls with size controls. Specification 4 includes both year controls and grade size controls. Errors were clustered on school campus groups.

Table 4.4: Structural model results

<table>
<thead>
<tr>
<th></th>
<th>Base model (1)</th>
<th>Year controls (2)</th>
<th>Size controls (3)</th>
<th>Full controls (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of other students</td>
<td>-14.76***</td>
<td>-9.61***</td>
<td>-10.95***</td>
<td>-8.45***</td>
</tr>
<tr>
<td></td>
<td>(2.43)</td>
<td>(2.35)</td>
<td>(1.60)</td>
<td>(2.00)</td>
</tr>
<tr>
<td>Classes taken with each student</td>
<td>0.16***</td>
<td>0.15***</td>
<td>0.14***</td>
<td>0.14***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.031)</td>
<td>(0.026)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Community Structure Found</td>
<td>0.077***</td>
<td>0.070***</td>
<td>0.074***</td>
<td>0.069***</td>
</tr>
<tr>
<td></td>
<td>(0.0079)</td>
<td>(0.0087)</td>
<td>(0.0075)</td>
<td>(0.0085)</td>
</tr>
</tbody>
</table>

Campus clustered standard errors in parentheses. Significance: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

Section 4.6: Discussion

The implementation of SLCs caused statistically significant changes in school structure, as expected. Students generally saw fewer students in classes each day and took more classes with the students they did see; there was also a significant increase in the level of community structure found by the community detection algorithms. Across the different specifications, it is apparent that while adding more controls reduces the magnitude of the effects, the effects of SLC implementation persist and are highly significant across all specifications. Including school size controls increases precision of the effect estimate, implying that they are beneficial to include,
but year-level controls increase the standard error, suggesting that the addition of year-level controls might be absorbing some effect of school structure presumably created by schools making changes prior to officially implementing SLCs. With both grade size and individual year controls, the number of students seen by each other student decreases by eight students when SLCs are implemented, and there is an increase in modularity of about 0.07, a very significant increase in structure.

**Practical significance**

Because little research has been done on the exact effects on a student of seeing fewer students each day in school, these effects are expressed both as a percentage, and as a proportion of the standard deviation present in the data. The decrease of eight students in number of students seen corresponds to a little more than 6% of the average number of 139 students, and around a quarter of the uncorrected standard deviation in number of students seen, not very significant. One reason that this effect might be so small is that the inflexibility of the high school master schedule and student prerequisites might preclude a further reduction in the average number of students in the average student’s classes (Johns, 2008b; D. Ready, et al., 2004): while SLCs might successfully schedule homogenous (one SLC) student groups in a couple of classes, the other classes might be even more heterogeneous. It is conceivable that, in the worst case, SLC implementation could successfully create communities where all students *only* take classes together but still increase the number of students seen each day.

The change in average number of classes shared with the average student of 0.14 is an increase of more than eleven percent of the base of 1.25 classes per student, and about half of the standard deviation. An alternate way to interpret this is to say that a student shares one more class within
each group of seven students. Overall, seeing the average of 139 students, each would share about nineteen more class periods with classmates with SLCs than before SLCs. A final way to interpret this is to count the number of additional periods shared across the average students’ class neighborhood. The base of 1.25 classes shared in an average of 139 students yields about 35 extra periods shared. (0.25 x 139). SLC implementation adds an additional 19, for a total of about 54 extra periods shared, a gain of more than 50%. This is a practically significant result. The effect size for this result is larger than that for the number of students seen each day, probably because even with the inflexibility of the master schedule, if a cohort of students now share several classes, this average will show it, even if the total number of students seen in classes each day stays the same or increases.

Finally, the increase in community structure (modularity) of 0.07 is a highly practically significant result. The community structure of schools which converted to SLCs was on average 0.304, with a standard deviation of 0.069. The increase of 0.07 is an increase of just over one standard deviation in community structure, a very practically significant increase in apparent structure. The size of this result is significant, and is probably because an increase in modularity is the first order desired effect of SLC implementation. While the average number of students seen or the average number of classes shared might stay the same, SLCs were supposed to create groups of students who took classes within their SLC – the very definition of modularity. While not perfectly segregated or “pure”, and while some community structure existed before, the implementation of SLCs created much more structurally well-defined communities of students inside schools.
Section 4.7: Conclusion

This chapter analyzed the changes in school structure resulting from the implementation of wall-to-wall SLCs. School structure was measured using a novel method from network analysis. The implementation of SLCs changed these school structural measures in practically and statistically significant ways. While results from Chapters 2 and 3 implied that groups of students in communities existed prior to SLC implementation, this analysis shows that school structure and the quantitative structural measures of the student experience also changed in the manner expected. SLCs are doing more than renaming communities that already exist (D. Ready, et al., 2004). The following chapter sheds light on which of these individual quantitative structural measures might be most important by presenting and discussing correlations between these measures and students’ individual academic outcomes.
Chapter 5: Correlations between school structure and academic outcomes

Section 5.1: Introduction

The previous chapter showed that the implementation of SLCs created changes in school structure; this chapter studies the correlations between structural measures and several academic outcomes. If SLCs are successfully simulating small schools, then this research should show a pattern of structural correlations similar to those schools (DeJong & Locker, 2006; M. Lee & Friedrich, 2007). While it is impossible to know whether structural changes cause changes in academic outcomes – it is reasonable to believe that SLC implementation could have changed both structure and academic outcomes independently – if a particular academic outcome is not correlated with any structural measures, it seems more plausible that any effects of SLC implementation on that academic outcome come from some other non-structural source (e.g., more or better professional development or more involved parents). Similarly, if this analysis determines that some structural measure is correlated with an academic outcome, and that correlation is invariant over the time frame of SLC implementation, then it suggests that the changes in that structural measure may possibly be a pathway for any changes to that academic outcome if the implementation of SLCs also changed that structural measure. On the contrary, if the correlation changed when SLCs were implemented, it may weaken this idea. It is important to reiterate that while the effects of SLCs on structure and academic outcomes can be estimated separately, it is impossible to know the direction or existence of causality between structure and academic outcomes, and therefore this chapter only presents correlations between the two.
This chapter describes the correlation of academic outcomes with several measures of school structure. First, the academic outcomes and structural measures are introduced. Next, the hypothesized correlations are described, and the models for testing them presented. Finally, results are shown for each academic outcome and structural measure and conclusions are drawn.

**Academic outcomes**
The academic outcomes studied in this chapter were chosen to cover a broad range of traditional and alternate indicators of student performance in school (Rumberger & Palardy, 2005), and span five domains of a student’s performance. They are as follows: Attendance, Discipline, GPA, Retention (as a proxy for dropping out), and Test Scores. Attendance is measured using the percent of school enrolled actually attended. Discipline is measured by the presence or absence of a suspension. GPA is measured in three ways: first, as an average; second, as whether a student earned a high GPA (> 3.3, ~90th percentile); finally, as whether a student earned a low GPA (< 0.7, ~10th percentile). Note the limitations of GPA described in Chapter 1. The measure of the retention domain is whether a student fails a grade (i.e., is retained) or disappears prior to advancing to the end of 12th grade (Miao & Haney, 2004). Finally, student test scores are measured by the z-score of the California English Language Arts annual test because it is given to all tested grades in high school, and for methodological concerns stated in Chapter 1.

**Structural measures**
The three structural measures used in this chapter are: the average size of a student’s community, the average number of classes taken with each connected student, and the fraction of connected students in the same community. The first two measures are described in Chapter 3. The last measure is the ratio of the number of unique connected students in the same calculated
community as the student to the total number of unique connected students, and captures the degree to which students take classes entirely with their “cohort” or community.

**Hypothesized correlations**

If SLCs work through their hypothesized pathway, changes in school structure related to having a smaller, more personalized learning environment should be positively correlated with beneficial academic outcomes (Cotton, 1996b, 2001; Leithwood & Jantzi, 2009; Raywid, 1996). For the three measures used in this chapter, within a school, students who share more classes with each other connected student, students who take classes with a higher percentage of students in their own structural *community*, and take classes with fewer students overall should all tend to have better academic outcomes, as suggested by the reviews of sociological literature (Allen & Steinberg, 2004; Johns, 2008a; V. E. Lee & Ready, 2006; Smerdon & Cohen, 2009).

**Section 5.2: Model**

This analysis explores correlations between academic outcomes and structural measures for all schools and years. Because an exploratory analysis indicated these correlations were different for ninth and tenth graders compared to eleventh and twelfth graders and that the correlations changed after the implementation of SLCs, this analysis explores differences in correlations between the upper and lower grades and, for those schools that implemented SLCs in the years of study, before and after the implementation of SLCs.

Two similar models are used for this analysis. The dependent variable in both models is the academic outcome. The independent variable of interest is the structural measure; its coefficient is estimated separately for the lower grades and upper grades and compared across the grade levels. Fixed-effect type controls are included for the following factors that might influence the
structural measures: year-based variation in correlation, school-grade variation in outcomes, year-grade variation in outcomes, and variation in outcomes and correlations based on the number of classes students take compared to the number the school allows. Because the focus of this analysis is on the raw, pooled correlations between uncorrected academic outcomes and structural measures not conditioned on a theoretical “average” student, and the exact interpretation of the counterfactual would be problematic anyway when including non-structural student-level factors, no controls for demographics or other non-structural student-level factors are desired or included in the model. Instead, the model compares each school-grade against itself, and each student is compared to other students who take the same relative number of classes, leaving in situ all other influences on correlations based on students’ choices within those classes, which SLCs affect. All errors are clustered by school. The base model is estimated over all schools and differences in correlations across the grade levels are calculated.

**Equation 5.1: Model specification 1: Base model**

\[
Y_{igt} = \delta_{a} I(Grades(a))S + \beta_{t}S + \beta_{g}X_{1igt}S + \beta_{s}X_{2igt} + \alpha_{sg} + \alpha_{gt} + \mu_{ist}
\]

\[
\mu_{ist} = \xi_{s} + \epsilon_{it}
\]

- \(Y = \) student level academic outcome
- \(S = \) structural measure
- \(X = \) vector of control variables
- \(s = \) school
- \(g = \) grade
- \(t = \) time
- \(i = \) student

School-grade, year-grade level fixed effects; year coefficient fixed effects

Errors clustered by school

The second model is similar to the first. The dependent variable is the academic outcome. The independent variable is the structural measure; its coefficient is estimated separately for grades nine and ten combined, and compared to grades eleven and twelve combined. For the second model, the coefficient is also estimated for these groups across the implementation of SLCs for a total of four correlations: pre-SLC ninth and tenth grades, pre-SLC eleventh and twelfth grades, post-SLC ninth and tenth grades, and post-SLC eleventh and twelfth grades. Fixed-effect type
controls are included for the following factors: year-based shared variation in correlation, school-
grade variation in outcomes and correlations, year-grade variation in outcomes, and variation in
outcomes and correlations based on the number of classes students take compared to the number
the school allows. Similar to the base model, correlations are estimated unconditional on student-
level variables other than the relative number of classes taken. All errors are clustered by school.
This second, extended model is estimated only over the forty campuses that implemented SLCs
in the time frame studied. For this model, differences between all combinations of grade levels
and SLC implementation status are calculated.

**Equation 5.2: Model specification 2: Including SLC implementation**

\[
Y_{igt} = \delta_{a,b} + (Grades(a), SLC(b))S + \beta_1S + \beta_2X_{1igt}S + \beta_3X_{2igt}S + \alpha_{igt} + \alpha_{igt} + \mu_{igt}
\]

\[
\mu_{igt} = \varepsilon_x + \varepsilon_u
\]

**Analysis limitations**

Since the purpose of this chapter is to provide a qualitative view of the pattern of correlations
between outcomes, the analyses and discussion in this chapter do not focus on interpreting the
particular correlation coefficient level, only the sign, relative magnitude, and statistical
significance. Standard errors are provided for the reader, but no correction for multiple
comparisons is made (Cohen, 1994; Rothman, 1990; Thompson, 1996).

**Model limitations**

These two models do not provide causal estimates of relationships between individual student
structural measures and academic outcomes; interpretation of results from this model must
consider that students, counselors, or other factors might cause simultaneous changes in both the
independent variable (network measure) and the dependent variables (academic outcome), and therefore completely explain or influence the correlation found. While the second model does provide before and after estimates of the change in correlation across SLC implementation, controlling for prior level and the effects of random time fluctuations, it is also non-causal; the model is completely agnostic of the all non-structural factors that can also influence the correlations; additionally while it does control for the relative number of classes taken, if the implementation of SLCs simultaneously changes the expected number of classes, the model results will be biased. Additionally, for simplicity, a linear probability model is used to estimate correlations between binary indicator variables.

**Section 5.3: Results**

**Attendance**

This portion of analysis examines correlations between structural measures and attendance. This analysis used the percent of school days attended as the outcomes variable. An analysis using a binary truancy indicator (absent > 20%) was conducted with similar qualitative results, and is therefore omitted. Table 5.1 shows correlations between the percent of schools days attended and the three different structural measures. The overall correlations using the first model are shown first, followed by results from the second model.
Table 5.1: Correlations with percent of schools days attended

<table>
<thead>
<tr>
<th></th>
<th>Fraction of connected students in same community (SG)</th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All Schools</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) 9th-10th grade</td>
<td>0.00094</td>
<td>0.00053*</td>
<td>0.10832</td>
</tr>
<tr>
<td></td>
<td>(0.00925)</td>
<td>(0.00023)</td>
<td>(0.05651)</td>
</tr>
<tr>
<td>(b) 11th-12th grade</td>
<td>-0.0112</td>
<td>0.00032</td>
<td>0.09032</td>
</tr>
<tr>
<td></td>
<td>(0.00964)</td>
<td>(0.00024)</td>
<td>(0.05648)</td>
</tr>
<tr>
<td>(a-b) Difference between grades</td>
<td>0.01214</td>
<td>0.00021**</td>
<td>0.018**</td>
</tr>
<tr>
<td></td>
<td>(0.00613)</td>
<td>(0.00003)</td>
<td>(0.00283)</td>
</tr>
<tr>
<td><strong>Schools eventually implementing SLCs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) Pre-SLC 9th-10th grade</td>
<td>-0.01289</td>
<td>0.00064*</td>
<td>0.15045**</td>
</tr>
<tr>
<td></td>
<td>(0.01175)</td>
<td>(0.00031)</td>
<td>(0.05318)</td>
</tr>
<tr>
<td>(d) Pre-SLC 11th-12th grade</td>
<td>-0.01917</td>
<td>0.0005</td>
<td>0.1428*</td>
</tr>
<tr>
<td></td>
<td>(0.01166)</td>
<td>(0.00032)</td>
<td>(0.05278)</td>
</tr>
<tr>
<td>(c-d) Difference between grades</td>
<td>0.00629</td>
<td>0.00014**</td>
<td>0.00765</td>
</tr>
<tr>
<td></td>
<td>(0.0103)</td>
<td>(0.00005)</td>
<td>(0.00436)</td>
</tr>
<tr>
<td>(e) Post-SLC 9th-10th grade</td>
<td>-0.01725</td>
<td>0.00065*</td>
<td>0.1483**</td>
</tr>
<tr>
<td></td>
<td>(0.01411)</td>
<td>(0.0003)</td>
<td>(0.05451)</td>
</tr>
<tr>
<td>(f) Post-SLC 11th-12th grade</td>
<td>-0.02676*</td>
<td>0.00049</td>
<td>0.14084*</td>
</tr>
<tr>
<td></td>
<td>(0.01228)</td>
<td>(0.00032)</td>
<td>(0.05359)</td>
</tr>
<tr>
<td>(f-e) Difference between grades</td>
<td>0.00951</td>
<td>0.00016**</td>
<td>0.00745*</td>
</tr>
<tr>
<td></td>
<td>(0.00813)</td>
<td>(0.00006)</td>
<td>(0.00341)</td>
</tr>
<tr>
<td><strong>Before and after SLC grade level comparisons</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e-c) 9th-10th grade</td>
<td>-0.00436</td>
<td>0.00001</td>
<td>-0.00215</td>
</tr>
<tr>
<td></td>
<td>(0.00664)</td>
<td>(0.00003)</td>
<td>(0.00334)</td>
</tr>
<tr>
<td>(f-d) 11th-12th grade</td>
<td>-0.00759</td>
<td>-0.00001</td>
<td>-0.00196</td>
</tr>
<tr>
<td></td>
<td>(0.00695)</td>
<td>(0.00003)</td>
<td>(0.00344)</td>
</tr>
<tr>
<td>(e-c) &amp; (f-d) Joint test (avg. shown)</td>
<td>-0.00598</td>
<td>0</td>
<td>-0.00205</td>
</tr>
<tr>
<td></td>
<td>(0.0068)</td>
<td>(0.00003)</td>
<td>(0.00339)</td>
</tr>
</tbody>
</table>

*Campus clustered standard errors in parentheses. Significance: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

For attendance, the number of classes taken with each other student is of most importance.

Students taking more classes with each other connected student tend to have higher rates of attendance. The size of the student’s class neighborhood is also somewhat positively related to having better attendance, with the strength of that correlation being greater in the ninth and tenth grades. Finally, the fraction of students in the same community has a generally negative, non-significant correlation with attendance, with that correlation reaching statistical significance in
the post-SLC eleventh and twelfth grades. This might be due to the simultaneous increase in both community structure (as shown in Chapter 3) and the use of “passport” (e.g., AP, music, advanced math, sports, student government, etc.) classes for more advanced students (DeJong & Locker, 2006; Matthews & Kitchen, 2007; Price, et al., 2003). These results are consistent with the hypothesis that students with good attendance took “passport” classes in the upper grades after the implementation of SLCs.

In conclusion, students that take more classes with more students tend to come to school more often, especially in the ninth and tenth grades. This results support the correlations that Raywid (1996), Smerdon (2009), Wasley (2000), and James-Burdumy (2008) also found. The implementation of SLCs did not change this situation.

**Discipline**

This analysis centers on the correlation between student discipline and structural measures.

Discipline is measured by whether a student was suspended at all during the year. Table 5.2 shows the results of the three correlation analyses.
Table 5.2: Correlations with being suspended

<table>
<thead>
<tr>
<th></th>
<th>Fraction of connected students in same community (SG)</th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Schools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) 9th-10th grade</td>
<td>-0.05167* (0.02274)</td>
<td>-0.00007 (0.00036)</td>
<td>-0.21022** (0.0402)</td>
</tr>
<tr>
<td>(b) 11th-12th grade</td>
<td>-0.01367 (0.00977)</td>
<td>-0.00035** (0.00005)</td>
<td>-0.02875** (0.00659)</td>
</tr>
<tr>
<td>(a-b) Difference</td>
<td>-0.038 (0.02268)</td>
<td>0.00029 (0.00035)</td>
<td>-0.18147** (0.0389)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) Pre-SLC 9th-10th grade</td>
<td>-0.06549* (0.02772)</td>
<td>-0.00031 (0.00043)</td>
<td>-0.18899** (0.03725)</td>
</tr>
<tr>
<td>(d) Pre-SLC 11th-12th grade</td>
<td>-0.04017 (0.02614)</td>
<td>-0.00004 (0.00098)</td>
<td>-0.16462** (0.03574)</td>
</tr>
<tr>
<td>(c-d) Difference</td>
<td>-0.02532 (0.01666)</td>
<td>-0.00034** (0.00041)</td>
<td>-0.02437* (0.01184)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e) Post-SLC 9th-10th grade</td>
<td>-0.06067 (0.03084)</td>
<td>-0.00032 (0.00043)</td>
<td>-0.18498** (0.03945)</td>
</tr>
<tr>
<td>(f) Post-SLC 11th-12th grade</td>
<td>-0.0397 (0.02952)</td>
<td>0.00003 (0.0004)</td>
<td>-0.16528** (0.03737)</td>
</tr>
<tr>
<td>(f-e) Difference</td>
<td>-0.02097 (0.01337)</td>
<td>-0.00035** (0.0001)</td>
<td>-0.0197 (0.00979)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e-c) 9th-10th grade</td>
<td>0.00482 (0.01611)</td>
<td>-0.00001 (0.00007)</td>
<td>0.00402 (0.00733)</td>
</tr>
<tr>
<td>(f-d) 11th-12th grade</td>
<td>0.00047 (0.0102)</td>
<td>-0.00001 (0.00004)</td>
<td>-0.00066 (0.00432)</td>
</tr>
<tr>
<td>(e-c) &amp; (f-d) Joint test (avg. shown)</td>
<td>0.00265 (0.01315)</td>
<td>-0.00001 (0.00006)</td>
<td>0.00168 (0.00582)</td>
</tr>
</tbody>
</table>

Schools eventually implementing SLCs

<table>
<thead>
<tr>
<th></th>
<th>Number of students in same community (SG)</th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Schools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) 9th-10th grade</td>
<td>-0.06549* (0.02772)</td>
<td>-0.00031 (0.00043)</td>
<td>-0.18899** (0.03725)</td>
</tr>
<tr>
<td>(d) 11th-12th grade</td>
<td>-0.04017 (0.02614)</td>
<td>0.00004 (0.00098)</td>
<td>-0.16462** (0.03574)</td>
</tr>
<tr>
<td>(c-d) Difference</td>
<td>-0.02532 (0.01666)</td>
<td>-0.00034** (0.00041)</td>
<td>-0.02437* (0.01184)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e) Post-SLC 9th-10th grade</td>
<td>-0.06067 (0.03084)</td>
<td>-0.00032 (0.00043)</td>
<td>-0.18498** (0.03945)</td>
</tr>
<tr>
<td>(f) Post-SLC 11th-12th grade</td>
<td>-0.0397 (0.02952)</td>
<td>0.00003 (0.0004)</td>
<td>-0.16528** (0.03737)</td>
</tr>
<tr>
<td>(f-e) Difference</td>
<td>-0.02097 (0.01337)</td>
<td>-0.00035** (0.0001)</td>
<td>-0.0197 (0.00979)</td>
</tr>
</tbody>
</table>

Before and after SLC grade level comparisons

<table>
<thead>
<tr>
<th></th>
<th>Number of students in same community (SG)</th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Schools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e-c) 9th-10th grade</td>
<td>0.00482 (0.01611)</td>
<td>-0.00001 (0.00007)</td>
<td>0.00402 (0.00733)</td>
</tr>
<tr>
<td>(f-d) 11th-12th grade</td>
<td>0.00047 (0.0102)</td>
<td>-0.00001 (0.00004)</td>
<td>-0.00066 (0.00432)</td>
</tr>
<tr>
<td>(e-c) &amp; (f-d) Joint test (avg. shown)</td>
<td>0.00265 (0.01315)</td>
<td>-0.00001 (0.00006)</td>
<td>0.00168 (0.00582)</td>
</tr>
</tbody>
</table>

Campus clustered standard errors in parentheses. Significance: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

From this table, the most significant correlation is with the number of classes taken with each student. Students who share more classes with each other student tend to have a lower rate of suspensions. This effect is uniform across schools and grades. In general, students in grades nine and ten are less likely to be suspended when having more unique students in their class neighborhood than students in grades eleven and twelve. Finally, while taking more classes with students in their own community tended to be protective against being suspended, the implementation of SLC weakened this effect.
In conclusion, students who stay out of disciplinary trouble are taking more classes with each connected student; and in ninth and tenth grades, taking classes with more students overall. The implementation of SLCs did not change this relationship. This is consistent with many studies of SLCs and small schools predicting improvements in discipline (Bernstein, et al., 2008; Cotton, 2001; James-Burdumy, et al., 2008; Leithwood & Jantzi, 2009; Wasley, et al., 2000)

**GPA**

While analysis of GPA has its problems, as discussed in Chapter 1, a student’s grade point average is one of the most lasting elements of his or her permanent record. For this analysis, the students’ annual grade point average is correlated with structural measures to determine the kind of relationship the average student has between their grade point average and the structural measures existing in their course network. Table 5.3 shows the results of correlations between students’ annual GPA and their individual structural measures.
Table 5.3: Correlations with GPA

<table>
<thead>
<tr>
<th></th>
<th>Fraction of connected students in same community (SG)</th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Schools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) 9th-10th grade</td>
<td>0.02387</td>
<td>0.0021</td>
<td>1.18054**</td>
</tr>
<tr>
<td></td>
<td>(0.05236)</td>
<td>(0.0011)</td>
<td>(0.20224)</td>
</tr>
<tr>
<td>(b) 11th-12th grade</td>
<td>0.0135</td>
<td>-0.00111</td>
<td>1.12628**</td>
</tr>
<tr>
<td></td>
<td>(0.07152)</td>
<td>(0.0011)</td>
<td>(0.19685)</td>
</tr>
<tr>
<td>(a-b) Difference</td>
<td>0.01036</td>
<td>0.00321**</td>
<td>0.05426</td>
</tr>
<tr>
<td>between grades</td>
<td>(0.06352)</td>
<td>(0.00035)</td>
<td>(0.03499)</td>
</tr>
<tr>
<td>Schools eventually implementing SLCs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) Pre-SLC 9th-10th grade</td>
<td>-0.0597</td>
<td>0.00303*</td>
<td>1.2945**</td>
</tr>
<tr>
<td></td>
<td>(0.06947)</td>
<td>(0.00135)</td>
<td>(0.17727)</td>
</tr>
<tr>
<td>(d) Pre-SLC 11th-12th grade</td>
<td>-0.04641</td>
<td>0.00054</td>
<td>1.31611**</td>
</tr>
<tr>
<td></td>
<td>(0.07376)</td>
<td>(0.00135)</td>
<td>(0.1718)</td>
</tr>
<tr>
<td>(c-d) Difference</td>
<td>-0.01328</td>
<td>0.00249**</td>
<td>-0.02162</td>
</tr>
<tr>
<td>between grades</td>
<td>(0.08612)</td>
<td>(0.00051)</td>
<td>(0.0567)</td>
</tr>
<tr>
<td>(e) Post-SLC 9th-10th grade</td>
<td>-0.15806</td>
<td>0.00287*</td>
<td>1.24279**</td>
</tr>
<tr>
<td></td>
<td>(0.08549)</td>
<td>(0.00132)</td>
<td>(0.18218)</td>
</tr>
<tr>
<td>(f) Post-SLC 11th-12th grade</td>
<td>-0.18133*</td>
<td>-0.00001</td>
<td>1.25053**</td>
</tr>
<tr>
<td></td>
<td>(0.07224)</td>
<td>(0.00132)</td>
<td>(0.18129)</td>
</tr>
<tr>
<td>(f-e) Difference</td>
<td>0.02327</td>
<td>0.00288**</td>
<td>-0.00774</td>
</tr>
<tr>
<td>between grades</td>
<td>(0.07995)</td>
<td>(0.00052)</td>
<td>(0.04856)</td>
</tr>
<tr>
<td>Before and after SLC grade level comparisons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e-c) 9th-10th grade</td>
<td>-0.09837</td>
<td>-0.00016</td>
<td>-0.05171</td>
</tr>
<tr>
<td></td>
<td>(0.05405)</td>
<td>(0.00024)</td>
<td>(0.02709)</td>
</tr>
<tr>
<td>(f-d) 11th-12th grade</td>
<td>-0.13492*</td>
<td>-0.00055</td>
<td>-0.06559*</td>
</tr>
<tr>
<td></td>
<td>(0.05428)</td>
<td>(0.00027)</td>
<td>(0.02879)</td>
</tr>
<tr>
<td>(e-c) &amp; (f-d) Joint test (avg. shown)</td>
<td>-0.11664*</td>
<td>-0.00035</td>
<td>-0.05865</td>
</tr>
<tr>
<td></td>
<td>(0.05416)</td>
<td>(0.00026)</td>
<td>(0.02794)</td>
</tr>
</tbody>
</table>

Campus clustered standard errors in parentheses. Significance: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

The table of grade point average shows that, for all students, taking more classes with each connected student is highly correlated with having a higher grade point average that year.

Students that take more classes with students in their same community tend to have lower grade point averages, especially after implementation of SLCs in eleventh and twelfth grades – potential evidence of higher performing students taking “passport” classes outside of the standard SLC curriculum (Price, et al., 2003).
Finally, the sign of correlations between the number of unique connected students and grade point average was different between the lower and upper grades; in the lower grades, having more connected students was correlated with having a higher grade point average. Conversely, having more connected students was correlated with a lower grade point average in eleventh and twelfth grades. These two results are compatible with the hypothesis that advanced students in the lower grades take classes with their entering cohort as well as with a more advanced cohort (e.g., upper-class mathematics) (Matthews & Kitchen, 2007); when they transition to the upper grades, and presumably even more advanced classes, they are then taking smaller classes with only upperclassmen, while their less academically advanced upper grade peers are sharing larger classes with underclassmen. The implementation of SLCs increased this correlation in eleventh and twelfth grades. These results are also consistent with the hypothesis that students with higher GPAs on average tended to share more classes with each student they saw during the day across all grades and, in the upper grades after the implementation of SLCs, share more classes with students outside of the community to which they belonged (again, presumably through advanced “passport” classes) (Price, et al., 2003).

Because this analysis indicated that students with very high and very low annual grade point averages had a pattern of structural correlations qualitatively different than those of the average student, further analysis into those student types was conducted and those results follow.

**High GPA**
This analysis examines students who have earned a high GPA during the year. Students who earn very high GPAs (top 10%) are different than other students and may have different perceptions and patterns of correlations (Marcoulides, Heck, & Papanastasiou, 2005). A similar,
supplemental analysis was performed for these students, and the results are presented in Table 5.4.

<table>
<thead>
<tr>
<th></th>
<th>Fraction of connected students in same community (SG)</th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All Schools</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) 9th-10th grade</td>
<td>-0.02674**</td>
<td>-0.00061*</td>
<td>-0.00332</td>
</tr>
<tr>
<td></td>
<td>(0.00908)</td>
<td>(0.00026)</td>
<td>(0.05011)</td>
</tr>
<tr>
<td>(b) 11th-12th grade</td>
<td>0.02968</td>
<td>-0.00167**</td>
<td>0.0395</td>
</tr>
<tr>
<td></td>
<td>(0.01535)</td>
<td>(0.00027)</td>
<td>(0.04854)</td>
</tr>
<tr>
<td>(a-b) Difference between grades</td>
<td>-0.05642**</td>
<td>0.00106**</td>
<td>-0.04283**</td>
</tr>
<tr>
<td></td>
<td>(0.01396)</td>
<td>(0.0001)</td>
<td>(0.01304)</td>
</tr>
<tr>
<td><strong>Schools eventually implementing SLCs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) Pre-SLC 9th-10th grade</td>
<td>-0.04183**</td>
<td>-0.00033</td>
<td>-0.00395</td>
</tr>
<tr>
<td></td>
<td>(0.01046)</td>
<td>(0.00034)</td>
<td>(0.06202)</td>
</tr>
<tr>
<td>(d) Pre-SLC 11th-12th grade</td>
<td>0.0217</td>
<td>-0.00127**</td>
<td>0.03644</td>
</tr>
<tr>
<td></td>
<td>(0.01311)</td>
<td>(0.00034)</td>
<td>(0.06291)</td>
</tr>
<tr>
<td>(c-d) Difference between grades</td>
<td>-0.06353**</td>
<td>0.00094**</td>
<td>-0.04039*</td>
</tr>
<tr>
<td></td>
<td>(0.01468)</td>
<td>(0.00009)</td>
<td>(0.01611)</td>
</tr>
<tr>
<td>(e) Post-SLC 9th-10th grade</td>
<td>-0.06741**</td>
<td>-0.00036</td>
<td>-0.01453</td>
</tr>
<tr>
<td></td>
<td>(0.01579)</td>
<td>(0.00034)</td>
<td>(0.0637)</td>
</tr>
<tr>
<td>(f) Post-SLC 11th-12th grade</td>
<td>-0.02602*</td>
<td>-0.00143**</td>
<td>0.01663</td>
</tr>
<tr>
<td></td>
<td>(0.01253)</td>
<td>(0.00033)</td>
<td>(0.06477)</td>
</tr>
<tr>
<td>(f-e) Difference between grades</td>
<td>-0.04139**</td>
<td>0.00106**</td>
<td>-0.03117*</td>
</tr>
<tr>
<td></td>
<td>(0.01323)</td>
<td>(0.00011)</td>
<td>(0.01428)</td>
</tr>
<tr>
<td><strong>Before and after SLC grade level comparisons</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e-c) 9th-10th grade</td>
<td>-0.02558*</td>
<td>-0.00003</td>
<td>-0.01059</td>
</tr>
<tr>
<td></td>
<td>(0.01157)</td>
<td>(0.00005)</td>
<td>(0.00564)</td>
</tr>
<tr>
<td>(f-d) 11th-12th grade</td>
<td>-0.04773**</td>
<td>-0.00016*</td>
<td>-0.01981*</td>
</tr>
<tr>
<td></td>
<td>(0.01322)</td>
<td>(0.00007)</td>
<td>(0.00732)</td>
</tr>
<tr>
<td>(e-c) &amp; (f-d) Joint test (avg. shown)</td>
<td>-0.03666**</td>
<td>-0.0001</td>
<td>-0.0152*</td>
</tr>
<tr>
<td></td>
<td>(0.01239)</td>
<td>(0.00006)</td>
<td>(0.00648)</td>
</tr>
</tbody>
</table>

Campus clustered standard errors in parentheses. Significance: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

As this table shows, the type of structure that shows the most correlations for students with high GPAs is the fraction of connected students in the same community. Students with high GPAs tend to take more classes with students outside of their community in the lower grades. Prior to whole-school SLC implementation, students with high GPAs in the upper grades tended to cluster in their own communities. After SLC implementation, students with high GPAs tended to take classes outside of their community as well. Effectively, SLC implementation destroyed or
obscured the existing communities of high performers (Attewell, 2001; D. Ready, et al., 2004). A look at correlations with the number of unique connected students hints that students with high GPAs tended to take classes with fewer other students, especially in the upper grades. Finally, while correlations with the number of classes taken with each student were not significantly different from zero, there is some evidence that the relationship is different between the upper and lower grades. It appears more beneficial advanced students in the lower grades to have fewer classes taken with each student, and detrimental for the upper grades to have fewer classes with each student. This is supportive of the hypothesis that advanced students take “passport” classes early in their high school tenure, settling down into a cohort later (Oakes & Guiton, 1995; Useem, 1991).

Summarizing, students with high GPAs tended to take fewer classes with a smaller number and wider range of students in the lower grades, and a narrower range of students in the upper grades. The implementation of SLCs made it more beneficial to take classes with a wider range of students in the upper grades as well. All of these findings are consistent with the hypothesis that high performing students tend to take a set of classes with a broad range of students in the lower grades, and have a more cohort-like experience in the upper grades (Frank, et al., 2008; Matthews & Kitchen, 2007).

Low GPA
Students with very low grade point averages also have different correlations with structural measures than the average student. Table 5.5 shows the results from an analysis of those students with low annual GPAs, under 0.7. These are students who are earning mostly Ds and Fails.
Table 5.5: Correlations with having a low GPA (<0.7)

<table>
<thead>
<tr>
<th></th>
<th>Fraction of connected students in same community (SG)</th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Schools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) 9th-10th grade</td>
<td>-0.03475 (0.03288)</td>
<td>-0.00224** (0.00041)</td>
<td>-0.52839** (0.07306)</td>
</tr>
<tr>
<td>(b) 11th-12th grade</td>
<td>-0.00995 (0.03338)</td>
<td>-0.00184** (0.00041)</td>
<td>-0.46812** (0.07112)</td>
</tr>
<tr>
<td>(a-b) Difference between grades</td>
<td>-0.02481 (0.01668)</td>
<td>-0.0004** (0.00008)</td>
<td>-0.6027** (0.00791)</td>
</tr>
<tr>
<td>(c) Pre-SLC 9th-10th grade</td>
<td>0.01104 (0.03572)</td>
<td>-0.00242** (0.00047)</td>
<td>-0.57887** (0.09721)</td>
</tr>
<tr>
<td>(d) Pre-SLC 11th-12th grade</td>
<td>0.05088 (0.03733)</td>
<td>-0.00222** (0.00046)</td>
<td>-0.53161** (0.09575)</td>
</tr>
<tr>
<td>(c-d) Difference between grades</td>
<td>-0.03983 (0.02456)</td>
<td>-0.0002 (0.00012)</td>
<td>-0.04726** (0.01175)</td>
</tr>
<tr>
<td>(e) Post-SLC 9th-10th grade</td>
<td>0.02291 (0.03787)</td>
<td>-0.00242** (0.00045)</td>
<td>-0.57077** (0.09765)</td>
</tr>
<tr>
<td>(f) Post-SLC 11th-12th grade</td>
<td>0.05743 (0.03303)</td>
<td>-0.00218** (0.00046)</td>
<td>-0.52888** (0.0969)</td>
</tr>
<tr>
<td>(f-e) Difference between grades</td>
<td>-0.03451 (0.02301)</td>
<td>-0.00024 (0.00013)</td>
<td>-0.04188** (0.00894)</td>
</tr>
<tr>
<td>(e-c) 9th-10th grade</td>
<td>0.01187 (0.01388)</td>
<td>0</td>
<td>0.0081 (0.00748)</td>
</tr>
<tr>
<td>(f-d) 11th-12th grade</td>
<td>0.00655 (0.01028)</td>
<td>0.00004 (0.00004)</td>
<td>0.00273 (0.00484)</td>
</tr>
<tr>
<td>(e-c) &amp; (f-d) Joint test (avg. shown)</td>
<td>0.00921 (0.01208)</td>
<td>0.00002 (0.00006)</td>
<td>0.00541 (0.00616)</td>
</tr>
</tbody>
</table>

Schools eventually implementing SLCs

<table>
<thead>
<tr>
<th></th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c) Pre-SLC 9th-10th grade</td>
<td>-0.00242** (0.00047)</td>
<td>-0.57887** (0.09721)</td>
</tr>
<tr>
<td>(d) Pre-SLC 11th-12th grade</td>
<td>-0.00222** (0.00046)</td>
<td>-0.53161** (0.09575)</td>
</tr>
<tr>
<td>(c-d) Difference between grades</td>
<td>-0.0002 (0.00012)</td>
<td>-0.04726** (0.01175)</td>
</tr>
<tr>
<td>(e) Post-SLC 9th-10th grade</td>
<td>-0.00242** (0.00045)</td>
<td>-0.57077** (0.09765)</td>
</tr>
<tr>
<td>(f) Post-SLC 11th-12th grade</td>
<td>-0.00218** (0.00046)</td>
<td>-0.52888** (0.0969)</td>
</tr>
<tr>
<td>(f-e) Difference between grades</td>
<td>-0.00024 (0.00013)</td>
<td>-0.04188** (0.00894)</td>
</tr>
</tbody>
</table>

Before and after SLC grade level comparisons

<table>
<thead>
<tr>
<th></th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e-c) 9th-10th grade</td>
<td>0</td>
<td>0.0081 (0.00748)</td>
</tr>
<tr>
<td>(f-d) 11th-12th grade</td>
<td>0.00004 (0.00004)</td>
<td>0.00273 (0.00484)</td>
</tr>
<tr>
<td>(e-c) &amp; (f-d) Joint test (avg. shown)</td>
<td>0.00002 (0.00006)</td>
<td>0.00541 (0.00616)</td>
</tr>
</tbody>
</table>

Campus clustered standard errors in parentheses. Significance: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

While it is impossible to determine the reasons for these correlations, especially considering the simultaneity of grading scales (McMillan, 2001; Starch & Elliott, 1912, 1913a, 1913b), instruction, and student effort (Marsh, 1987), this analysis shows that an increased number of classes shared with each connected student is protective against earning a poor GPA (i.e., for at-risk students) (Capps & Maxwell, 1999; Leithwood & Jantzi, 2009). This correlation is large across all the analyses and is stable across SLC implementation. A larger amount of connected students is also somewhat protective against earning a low GPA. The fraction of connected students in the same community is not significantly correlated to the chance of earning a low GPA.
GPA, unlike the results for earning a high GPA. This analysis implies that students who earn low GPAs tend to take fewer classes with fewer students and were no more formed into groups than the average student, perhaps indicating a sense of disengagement, which is contra to what the research literature has indicated (Hirschfield & Gasper, 2010; Leithwood & Jantzi, 2009; Weiss, et al., 2010)

**Retention**
Retention is measured by an indicator of whether the student was retained in grade (failed to earn enough credits to promote) or disappeared from school prior to twelfth grade (dropped out), two highly correlated events (Roderick, 1994). This analysis shows the correlation between this failure indicator and structure measures in the year before the failure. Results are displayed in Table 5.6.
<table>
<thead>
<tr>
<th></th>
<th>Fraction of connected students in same community (SG)</th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All Schools</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) 9th-10th grade</td>
<td>-0.00679</td>
<td>0.00077*</td>
<td>-0.10473</td>
</tr>
<tr>
<td></td>
<td>(0.03354)</td>
<td>(0.00033)</td>
<td>(0.12346)</td>
</tr>
<tr>
<td>(b) 11th-12th grade</td>
<td>0.08009*</td>
<td>0.00095*</td>
<td>0.04606</td>
</tr>
<tr>
<td></td>
<td>(0.03932)</td>
<td>(0.00037)</td>
<td>(0.12915)</td>
</tr>
<tr>
<td>(a-b) Difference between grades</td>
<td>-0.08688**</td>
<td>-0.00018</td>
<td>-0.15079**</td>
</tr>
<tr>
<td></td>
<td>(0.02466)</td>
<td>(0.00014)</td>
<td>(0.01672)</td>
</tr>
<tr>
<td><strong>Schools eventually implementing SLCs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) Pre-SLC 9th-10th grade</td>
<td>0.03412</td>
<td>0.00046</td>
<td>-0.1496</td>
</tr>
<tr>
<td></td>
<td>(0.03492)</td>
<td>(0.00036)</td>
<td>(0.13379)</td>
</tr>
<tr>
<td>(d) Pre-SLC 11th-12th grade</td>
<td>0.03289</td>
<td>0.00066</td>
<td>-0.10318</td>
</tr>
<tr>
<td></td>
<td>(0.05327)</td>
<td>(0.00039)</td>
<td>(0.13514)</td>
</tr>
<tr>
<td>(c-d) Difference between grades</td>
<td>0.00123</td>
<td>-0.00021</td>
<td>-0.04642</td>
</tr>
<tr>
<td></td>
<td>(0.04125)</td>
<td>(0.0002)</td>
<td>(0.02554)</td>
</tr>
<tr>
<td>(e) Post-SLC 9th-10th grade</td>
<td>0.00827</td>
<td>0.00033</td>
<td>-0.16108</td>
</tr>
<tr>
<td></td>
<td>(0.04401)</td>
<td>(0.0004)</td>
<td>(0.13061)</td>
</tr>
<tr>
<td>(f) Post-SLC 11th-12th grade</td>
<td>0.14886*</td>
<td>0.00128**</td>
<td>-0.05781</td>
</tr>
<tr>
<td></td>
<td>(0.05898)</td>
<td>(0.00041)</td>
<td>(0.13453)</td>
</tr>
<tr>
<td>(f-e) Difference between grades</td>
<td>-0.14059**</td>
<td>-0.00095**</td>
<td>-0.10327**</td>
</tr>
<tr>
<td></td>
<td>(0.03624)</td>
<td>(0.00023)</td>
<td>(0.02255)</td>
</tr>
<tr>
<td><strong>Before and after SLC grade level comparisons</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e-c) 9th-10th grade</td>
<td>-0.02585</td>
<td>-0.00013</td>
<td>-0.01148</td>
</tr>
<tr>
<td></td>
<td>(0.02268)</td>
<td>(0.0001)</td>
<td>(0.01157)</td>
</tr>
<tr>
<td>(f-d) 11th-12th grade</td>
<td>0.11597**</td>
<td>0.00062**</td>
<td>0.04537**</td>
</tr>
<tr>
<td></td>
<td>(0.01598)</td>
<td>(0.00009)</td>
<td>(0.00796)</td>
</tr>
<tr>
<td>(e-c) &amp; (f-d) Joint test (avg. shown)</td>
<td>0.04506**</td>
<td>0.00025**</td>
<td>0.01694**</td>
</tr>
<tr>
<td></td>
<td>(0.01933)</td>
<td>(0.00009)</td>
<td>(0.00977)</td>
</tr>
</tbody>
</table>

*Campus clustered standard errors in parentheses. Significance: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.*

The pattern of results for failure is simultaneously more complex and less significant than that for other academic measures. For schools that implemented SLCs, nothing is significantly correlated with failing except for students in upper grades after implementation. In that case, students who take classes with many students inside one large community tend to be more likely to fail or dropout. For all schools (including non/late implementers) this pattern also partly holds for lower grades. It is possible that this is consistent with a district or school policy to place all students who need remediation into a large classroom together, as suggested by Rayyes (2009).
Students who avoid failure tend to see fewer unique students in their classes, and after SLC implementation, upper grade students who are most likely to fail are taking most of their classes in one larger community – a result also consistent with the hypothesis that advanced students take classes outside of the typical SLC class (Oakes & Guiton, 1995; Price, et al., 2003; Useem, 1991).

**Test scores**
This analysis of test scores only includes English test score levels. Math and science test scores are excluded as described in Chapter 1. In California, the same English test is given to each student in a particular grade. Unlike the other academic outcomes, state tests are not given to seniors, and therefore grade twelve is omitted. Corresponding exploratory analyses similar to those conducted for high and low GPA indicated that students scoring very well or very poorly on their tests shared a pattern of correlations similar to the average student, and are therefore omitted. Additionally, an analysis of correlations of structure and test score gains showed no significant correlations in any case, and therefore is also omitted. Table 5.7 shows correlations with English test score levels.
### Table 5.7: Correlations with reading test score levels

<table>
<thead>
<tr>
<th></th>
<th>Fraction of connected students in same community (SG)</th>
<th>Number of unique connected students</th>
<th>Average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All Schools</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) 9th-10th grade</td>
<td>-0.01728</td>
<td>0.00234</td>
<td>-0.60023</td>
</tr>
<tr>
<td></td>
<td>(0.13958)</td>
<td>(0.00214)</td>
<td>(0.43872)</td>
</tr>
<tr>
<td>(b) 11th grade</td>
<td>0.11518</td>
<td>-0.00247</td>
<td>-0.4361</td>
</tr>
<tr>
<td></td>
<td>(0.1632)</td>
<td>(0.0022)</td>
<td>(0.41702)</td>
</tr>
<tr>
<td>(a-b) Difference between grades</td>
<td>-0.13246</td>
<td>0.00481**</td>
<td>-0.16414</td>
</tr>
<tr>
<td></td>
<td>(0.07842)</td>
<td>(0.00067)</td>
<td>(0.08728)</td>
</tr>
<tr>
<td><strong>Schools eventually implementing SLCs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) Pre-SLC 9th-10th grade</td>
<td>0.0847</td>
<td>0.00147</td>
<td>-0.27265</td>
</tr>
<tr>
<td></td>
<td>(0.146)</td>
<td>(0.00294)</td>
<td>(0.63877)</td>
</tr>
<tr>
<td>(d) Pre-SLC 11th grade</td>
<td>0.41773*</td>
<td>-0.0039</td>
<td>0.0911</td>
</tr>
<tr>
<td></td>
<td>(0.18364)</td>
<td>(0.00289)</td>
<td>(0.58912)</td>
</tr>
<tr>
<td>(c-d) Difference between grades</td>
<td>-0.33303**</td>
<td>0.00537**</td>
<td>-0.36376**</td>
</tr>
<tr>
<td></td>
<td>(0.11581)</td>
<td>(0.00087)</td>
<td>(0.1171)</td>
</tr>
<tr>
<td>(e) Post-SLC 9th-10th grade</td>
<td>-0.07505</td>
<td>0.00134</td>
<td>-0.30483</td>
</tr>
<tr>
<td></td>
<td>(0.16748)</td>
<td>(0.00294)</td>
<td>(0.65128)</td>
</tr>
<tr>
<td>(f) Post-SLC 11th grade</td>
<td>0.08589</td>
<td>-0.00494</td>
<td>-0.05839</td>
</tr>
<tr>
<td></td>
<td>(0.17988)</td>
<td>(0.00286)</td>
<td>(0.61465)</td>
</tr>
<tr>
<td>(f-e) Difference between grades</td>
<td>-0.16094**</td>
<td>0.00628**</td>
<td>-0.24644**</td>
</tr>
<tr>
<td></td>
<td>(0.08522)</td>
<td>(0.00083)</td>
<td>(0.10141)</td>
</tr>
<tr>
<td><strong>Before and after SLC grade level comparisons</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e-c) 9th-10th grade</td>
<td>-0.15974</td>
<td>-0.00014</td>
<td>-0.03218</td>
</tr>
<tr>
<td></td>
<td>(0.09572)</td>
<td>(0.00041)</td>
<td>(0.04747)</td>
</tr>
<tr>
<td>(f-d) 11th grade</td>
<td>-0.33183*</td>
<td>-0.00104</td>
<td>-0.14949*</td>
</tr>
<tr>
<td></td>
<td>(0.13674)</td>
<td>(0.00057)</td>
<td>(0.06788)</td>
</tr>
<tr>
<td>(e-c) &amp; (f-d) Joint test (avg. shown)</td>
<td>-0.24579</td>
<td>-0.00059*</td>
<td>-0.09084*</td>
</tr>
<tr>
<td></td>
<td>(0.11623)</td>
<td>(0.00049)</td>
<td>(0.05768)</td>
</tr>
</tbody>
</table>

Campus clustered standard errors in parentheses. Significance: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

Foreshadowing things to come, like James-Burdumy (2008), Shear (2008), and Bernstein (2008), not many structural measures have significant correlations with test scores (none affected test score gains). Only the eleventh grade fraction of connected students in the same community correlation is significantly different from zero, and then only before SLC implementation for those schools that eventually implemented SLCs. However, almost all categories of comparisons have a significant difference in correlations between the lower and upper grades. In all such cases, these differences are consistent with the hypothesis that higher scoring students tend to take a broad range of classes with a broad range of students in the lower grades; in the upper
grades, they tend to take classes with a smaller, more focused, narrow group of students (Oakes & Guiton, 1995). One more thing stands out here: pre-SLC implementation, only high scoring students were grouped and took classes in a fairly coherent community of students like Ready (2004) suggested; post-SLC implementation, all students were grouped in communities, and the fraction of connected students in the same community was no longer significantly correlated with test performance. These results further support the finding of Chapter 3: the implementation of SLCs significantly changed community structure.

Summarizing these results, students with high test scores tended to take classes with a broad range of students in the lower grades, while focusing into smaller communities and classes in the upper grades, as found by Oakes (1995). By grouping all students into communities, SLC implementation reduced the correlation between being in a community and test scores, especially in eleventh grade – a result consistent with the hypothesis that higher achieving students were grouped to a greater extent than other students before SLC implementation (it is unclear whether SLCs reduced the grouping of higher achieving students, increased the grouping of other students, or both), and that advanced students tend to take a broad range of classes with other students in the lower grades and enter a smaller cohort in the upper grades. This finding supports the hypothesis that because some selective non-whole-school SLCs were present in schools prior to full implementation, and entry into those SLCs was voluntary (Price, et al., 2003), higher-performing students, parents, and/or teachers had selected into these SLCs (Delany, 1991; Neild, 2005; D. D. Ready & Lee, 2008). This also suggests that an abundance of caution should be used when interpreting results from studies without controls for differences in student, parent, and teacher motivation, like Bernstein (2008) and Stern (2008), among others. Finally, this raises a
question about high performing students: if they were grouped before, and they are not grouped now, will anything change about their education outcomes? Marsh (1987) indicates that they might be better off being relatively bigger fish in their new ponds, but Bryk (1988) and Gamoran (1992) present contrary evidence. It is a topic for further research.

**Section 5.4: Summary of results**

This chapter reviews many different correlations between students’ academic outcomes and structural measures in the school. To help visualize the pattern of results, Table 5.8 summarizes these analyses. Note that in this table, the correlation sign for negative outcomes is inverted, contrary to elsewhere in this chapter — allowing for a consistent interpretation of the “+” and “-” symbols.

<table>
<thead>
<tr>
<th>Structural Measure</th>
<th>Greater Fraction of connected students in same community (SG)</th>
<th>Greater number of unique connected students</th>
<th>Greater average number of classes taken with each connected student</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade Level</strong></td>
<td>9th/10th 11th/12th</td>
<td>9th/10th 11th/12th</td>
<td>9th/10th 11th/12th</td>
</tr>
<tr>
<td>Attendance</td>
<td>o/o/o o/o/-</td>
<td>+/-/+ o/o/o</td>
<td>o/+/+ o/+/+</td>
</tr>
<tr>
<td>(-)Suspensions</td>
<td>+/-/+ o/o/o</td>
<td>o/o/o o/o/o</td>
<td>+/-/+ +/-/+</td>
</tr>
<tr>
<td>Average GPA</td>
<td>o/o/o o/o/-</td>
<td>o/+/+ o/o/o</td>
<td>+/-/+ +/-/+</td>
</tr>
<tr>
<td>High GPA</td>
<td>+/-/- o/o/-</td>
<td>+/-/- o/o/-</td>
<td>o/o/o o/o/o</td>
</tr>
<tr>
<td>(-)Low GPA</td>
<td>o/o/o o/o/o</td>
<td>+/-/+ +/-/+</td>
<td>+/-/+ +/-/+</td>
</tr>
<tr>
<td>(-)Failure</td>
<td>o/o/o -/o/-</td>
<td>-/o/o -/o/-</td>
<td>o/o/o o/o/o</td>
</tr>
<tr>
<td>Test scores</td>
<td>o/o/o o/+/-</td>
<td>o/o/o o/o/o</td>
<td>o/o/o o/o/o</td>
</tr>
<tr>
<td>Gain test scores</td>
<td>o/o/o o/o/o</td>
<td>o/o/o o/o/o</td>
<td>o/o/o o/o/o</td>
</tr>
</tbody>
</table>

Key: The symbols +, -, o signify academically positive, negative, and non-significant correlations, respectively. The symbol order within each column gives correlations for overall student body/pre-SLC implementation students/post-SLC implementation students, respectively. Correlations for Discipline (suspending), Low GPA, and Failure are inverted so that + is always academically beneficial.

Taking each structural measure sequentially, it is clear that a greater fraction of connected students in a community yields the most interesting results. The lower grades appear split into two groups of students: in the first, average group, having a more coherent community was
correlated with fewer suspensions, a beneficial disciplinary effect predicted by Hirschfield (2010); in the second, advanced group, the opposite correlation came into play, having a far-flung set of connected students was a mark of very high grades, a result suggested by sorting mechanisms in high school (Oakes & Guiton, 1995). The upper grades, meanwhile, corroborate the hypothesis that high ability groups (as also shown in Chapter 3) existed prior to whole-school SLC implementation (D. Ready, et al., 2004), and that the implementation of SLCs essentially diluted them such that those students then needed to take classes outside of their community – a result potentially due to the direction to avoid grouping high ability students into one SLC (Rayyes & Barela, 2009).

When reviewing results for the number of students connected, it seems that students with better outcomes took classes with more unique students in the lower grades, and then took classes with fewer unique students in the upper grades, a trend that remained across the implementation of SLCs, and again provide a demonstration of the “high school sorting machine” in action (Delany, 1991; Oakes & Guiton, 1995). Finally, sharing more classes with each connected student was always correlated with better outcomes. That finding in particular and these findings in general, are broadly consistent with the sociological justifications for small schools and SLCs and the literature on high school processes (Cotton, 2001; Leithwood & Jantzi, 2009; Levine, 2010; Oakes & Guiton, 1995; Rayyes & Barela, 2009).

**Section 5.5: Discussion**

The totality of these results suggests that students who have higher average academic outcomes tend to share fewer classes with a large number of unique students as an underclassman. In the
upper grades, they tend to share more classes with each student in a smaller cohort of students, while taking classes outside of their SLCs.

Additionally, these correlation analyses show that, as might be expected, and as shown in Chapter 3, even prior to the implementation of whole-school SLCs, coherent communities of students who took classes together already existed. These communities tended to be composed of students with (sometimes much) better academic outcomes than the average student. After SLC implementation, all students were formed into communities, and taking classes with students inside a community was no longer correlated with better outcomes, contrary to what was hypothesized and what existed before. Indeed, there is good evidence that students with better outcomes take “passport” classes, classes outside of their community, in upper grades (Price, et al., 2003; Rayyes & Barela, 2009). The implementation of SLCs appears to have reduced the grouping of students with better academic outcomes.

**Section 5.6: Conclusion**

The previous chapter showed that the implementation of SLCs had effects on school structure. After implementation, the number of classes shared with each connected student increased, while the average number of unique connected students decreased. While not causal, results from the present correlation analyses suggest that those two structural changes, especially the former, may be pathways to better academic outcomes for students, as suggested by Cotton (2001). Students who share more classes with other students have a greater chance of being friends with someone with whom they share several classes (Babcock, 2008; Jackson & Rogers, 2007). Friendship with another person who takes most of the same classes likely provides a higher return on time invested, encouraging students to form stronger connections within the protective environment of
school instead of forming riskier connections with people outside of school (Weiss, et al., 2010), possibly reducing adolescent delinquency (Hirschfield & Gasper, 2010).

Unfortunately, this analysis, in a result also shared by Bernstein (2008) and Shear (2008), failed to find structural correlations with test score levels or gains, though the results for students’ softer academic outcomes (attendance, discipline, GPA, and grade retention) suggest that the implementation of SLCs may have beneficial effects on students via a structural change pathway where students share more classes together in smaller cohorts (Cotton, 1996b, 2001; Raywid, 1996). Interestingly, the results for the percent of classes taken within the students calculated community suggest that though the implementation of SLCs did have a practically significant effect on the level of community structure, modularity (Chapter 4), instead of the percentage of classes taken with community members (stronger cohort) providing a potential pathway for change in academic outcomes, the community creating portion of SLC implementation simply reshuffled groups of students – the instability of the correlations with the percentage of classes taken with other students in the same community seem to reflect scheduling practices (Neubig, 2006), more than a stable mechanism to support feelings of connectedness (McNeely, et al., 2002), or produce better academic outcomes (Cotton, 1996b).
Chapter 6: Effect of SLC on student outcomes

Section 6.1: Introduction

This chapter’s aim is to discuss the effect of the implementation of whole-school small learning communities on academic outcomes. The Board of Education hoped to change many academic outcomes and attributes of schools for the better by implementing whole-school or “wall-to-wall” SLCs across all high schools, including the following (Price, et al., 2003; Rayyes & Barela, 2009; Tokofsky & Lauritzen, 2004):

- Have a distinct and compelling vision, mission, goals, objectives, and clearly recognizable identity
- Have between 350-500 students, and an inclusive admission policy.
- Help close the achievement gap
- Provide coherent, rigorous standards-based curriculum
- Focus on the needs of students within the SLC
- Create a spirit of collaboration
- Increase the numbers of students graduating
- Provide a personalized learning environment where every child is well known by a group of educators

This causal analysis focuses on determining the effect of SLC implementation on attendance, discipline, GPA, grade retention, and test scores. The hypothesis is that SLCs can be expected to increase attendance, GPA, and test scores while reducing grade retention and discipline issues, as suggested by Raywid (1996) for SLCs and Leithwood & Jantzi (2009) for small schools.

Section 6.2: The model

The model used for the analysis is a comparative interrupted time series model (Bloom, 2003; Bloom & Riccio, 2005; Shadish, et al., 2002). The variation in implementation date both within and across schools is used to estimate the effect of implementing SLCs across an entire grade.
Because schools often implemented SLCs in some grades before others (Rayyes & Barela, 2009), for example, starting with ninth grade academies and opening a new grade each year, each grade is considered separately. Each school and grade is used as its own control. Therefore, schools that did not experience a change in SLC implementation status are largely excluded from the model. Outcomes are either averaged at the school-year-grade level or left alone. The base specification used to model the effects of SLCs is given in Equation 6.1. The coefficient on a grade by year by school matrix of indicators of SLC implementation is interpreted as the causal effect of SLC implementation on that outcome. Robust errors are clustered by school to account for any intra-school correlations, such as school level data entry errors, and the fact that students are generally nested within schools (Wooldridge, 2002, 2003).

**Equation 6.1: Model specification 1: Base model**

\[
Y_{sgr} = \delta I(SLC)_{sgr} + \alpha_s + \alpha_g + \mu_{ist}
\]

- \(\mu_{ist} = \epsilon_s + \epsilon_{ist}\)
- \(Y = \text{student or school level outcomes}\)
- \(s = \text{school}\)
- \(g = \text{grade}\)
- \(t = \text{time}\)
- \(i = \text{students}\)
- School-grade, year fixed effects
- Errors clustered by school

An additional model specification with a range of student-level controls is included (Equation 6.2). The controls used are: gifted status, special education status, ethnicity, English language learner status, language spoken at home, free/reduced price lunch, gender, habitual truant, and whether student was suspended. When one of these is the outcome measured, it is removed from the set of covariates.
Equation 6.2: Model specification 2: Additional covariates

\[ Y_{ist} = \delta I(SLC)_{sgt} + \beta X_{ist} + \alpha_s + \alpha_g + \alpha_t + \mu_{ist} \]

\[ \mu_{ist} = \varepsilon_i + \varepsilon_{ist} \]

\( Y = \) student or school level outcomes
\( X = \) student or school level covariates
\( s = \) school
\( g = \) grade
\( t = \) time
\( i = \) students

School-grade, year fixed effects
Errors clustered by school

Section 6.3: Results for academic outcomes

Attendance

School attendance is important for students as well as school districts: in California, students that do not attend school also cause the district to lose state funding. While this analysis does not attempt to tease out the particular reasons for any effects, the implementation of SLCs may improve school attendance in several common sense ways: Feeling more connected may encourage students to attend school (McNeely, et al., 2002); being a member of a SLC might make students happier and take fewer “sick” days (Weitzman et al., 1986); SLC counselors might do a better job following up on severely truant students, and encourage them to attend school as in Allen (2001); finally, the increased personalization of SLCs might encourage students to attend by persuading them they have a bright future if they come to school (Cotton, 2001).

Two attendance measures are used for this analysis. The first measure is an individual measure of the percentage of days that a student attends school. The second measure is the average number of students at the school who are truant, absent 20% of the time or more. Table 6.1 shows the effect of SLCs on these measures.
Table 6.1: Effect of SLCs on attendance

<table>
<thead>
<tr>
<th></th>
<th>Fraction of school attended, base model</th>
<th>Fraction of school attended, full model</th>
<th>Average truant, base model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of SLC</td>
<td>-0.00108 (0.00211)</td>
<td>-0.00200 (0.00180)</td>
<td>-0.00390 (0.00619)</td>
</tr>
<tr>
<td>implementation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses. Significance: $+ = p < 0.10$, $* = p < 0.05$, $** = p < 0.01$, $*** = p < 0.001$.

As these results show, the implementation of SLCs had no statistically significant effect on students’ school attendance or truancy. This result is contrary to the associations from Chapter 5 as well as other research on small schools and SLCs such as Raywid (1996), Smerdon (2009), Wasley (2000), and James-Burdumy (2008) who all found positive correlations between students SLCs and attendance. However, as previously mentioned in this document, these studies generally did not control for student selection on unobservables as well as the current research, and might be further explained by the relatively high level, seriously heteroskedastic nature, and use of conservative error structure for analysis of attendance (Wooldridge, 2006).

**Discipline**

The analysis now turns to a discussion of student discipline as measured by the number of students suspended from school. Reiterating, this analysis is not designed to provide guidance as to what factors may contribute to any SLC effects on student suspensions. Several common sense potential hypotheses to help aid interpretation are provided: the greater feeling of connection can reduce the urge of students to act out (Hirschfield & Gasper, 2010; McNeely, et al., 2002); the fact that a student sees fewer students each day can increase their potential to work out issues on their own (Peterson & Skiba, 2001; Wasley, et al., 2000); students greater engagement and teachers’ better understanding of individual students may encourage them to pursue other means of discipline instead of suspension (V. E. Lee & Smith, 1995; Rosenfeld, Richman, & Bowen, 2000; Weiss, et al., 2010); finally, students might be sent to their SLC counselor for discipline instead of a dean (Rayyes & Barela, 2009), a result that might be less
likely to end in suspension (Klonsky, 2002; Sugai, Sprague, Horner, & Walker, 2000). Clearly, any improved student behavior supports a better learning environment (Dupper & Meyer-Adams, 2002).

This analysis uses two ways to measure discipline: number of students suspended from school. The first measure is the average fraction of students suspended each year. The second measure is the difference in probability (marginal effect) of a student being suspended. Table 6.2 shows these results.

<table>
<thead>
<tr>
<th>Effect of SLC implementation</th>
<th>Average fraction of students suspended, base model</th>
<th>Probability of being suspended, full model</th>
<th>Probability of being suspended, base model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.0148*</td>
<td>-0.0105*</td>
<td>-0.0110*</td>
</tr>
<tr>
<td></td>
<td>(0.00741)</td>
<td>(0.00426)</td>
<td>(0.00435)</td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses. Significance: + = p < 0.10, * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

The results of this analysis show that SLCs have improved discipline outcomes on average. The implementation of SLCs caused about a one and one-half percent drop in the number of students suspended each year. Additionally, the individual probability of a student being suspended dropped by a bit more than one percent, controlling for their other attributes. While it is unclear whether this improvement is through student level improvement in behavior or teacher- or school-level changes in disciplinary responses or procedures, a one percent drop is very practically significant given that the baseline probability of suspension in a given year is less than 10%. This positive effect somewhat agrees with the correlations from Chapter 5, and also agrees with nearly all previous studies on small schools and SLCs which found positive effects on suspensions, feelings of safety, or incidents of school violence (Bernstein, et al., 2008; Cotton, 1996a, 2001; James-Burdumy, et al., 2008; Leithwood & Jantzi, 2009; D. Ready, et al., 2004; Smerdon & Cohen, 2009; Wasley, et al., 2000).
GPA
While somewhat problematic to interpret causally, a student’s grade point average is one of the important and lasting elements of their high school record (Farrell Jr, Sapp, Johnson Jr, & Pollard, 1994). It determines what type of college they attend, and possibly even think of attending (Alwin & Otto, 1977). Long after their attendance and disciplinary records are forgotten, their high school GPA appears, on form after application form, for scholarships, (Cabrera, Nora, & Castaneda, 1992), and college admissions (Hurtado, Inkelas, Briggs, & Rhee, 1997; Nelson, 1972). If students perform better, and teachers grade accordingly, or even if, as found by Starch (1912) and McMillan (2001), teachers in a “smaller school” simply grade relatively easier, the implementation of SLCs can improve the students’ recorded GPA, remove barriers to attainment, and possibly their financial and academic lives well into the future (Hurtado, et al., 1997).

This analysis uses three measures of GPA. The first two are indicators of whether a student earned a high or low annual GPA (>3.3 and <0.7, respectively). These indicators are measured both as the probability of earning that level of GPA, as well as the aggregate fraction of students earning that GPA. The last measure is the simple annual grade point average. SLCs could possibly reduce the numbers of students failing nearly all classes, and increase the numbers of students earning mostly A’s and B’s, even if they do not affect the overall average. Table 6.3 shows the results of this analysis.
Table 6.3: Effect of SLCs on grades

<table>
<thead>
<tr>
<th>SLC Implementation</th>
<th>Probability of high GPA</th>
<th>Average fraction of students earning high GPA</th>
<th>Probability of low GPA</th>
<th>Average fraction earning low GPA</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Base model)</td>
<td>0.00400 (0.00257)</td>
<td>0.000423 (0.00453)</td>
<td>-0.00596* (0.00284)</td>
<td>-0.00720 (0.00469)</td>
<td>0.0176</td>
</tr>
<tr>
<td>(Full model)</td>
<td>0.00422+ (0.00255)</td>
<td>n/a</td>
<td>-0.00564* (0.00241)</td>
<td>n/a</td>
<td>0.0148</td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses. Significance: + = p < 0.10, * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

The results of these analyses indicate that SLCs do have a slight effect on students’ GPA. SLCs cause a mild, nearly statistically significant increase in the probability that a student would earn a high annual GPA controlling for their other characteristics. SLCs additionally cause a statistically significant drop of slightly more than one-half of a percent in the individual probability of earning a low GPA, after controlling for other student level covariates. Finally, the implementation of SLCs results in an insignificant increase in the annual average GPA. As described in this analysis’ preamble, earning a higher GPA may have beneficial effects on students’ futures that persist for a long time (Alwin & Otto, 1977; Attewell, 2001; Bryk & Driscoll, 1988; Marsh, 1987; Marsh & Parker, 1984). Note however, as described previously in Chapter 1, a higher GPA does not necessarily mean students are learning more; this analysis is unable to isolate the effect of SLCs from the effect of differences in teacher relative grading (i.e., easier grading curve) (McMillan, 2001). Still, whether the improvement in GPA is from actual learning or easier grading, the implementation of SLCs improves the permanent records and academic self-concept of lower-achieving, “at-risk” students (Farrell Jr, et al., 1994; Hurtado, et al., 1997). Few other studies of small schools or SLCs specifically mentioned GPA, often lumping it in with “academic performance” (Cotton, 2001). However, the one study that did, (Wasley, et al., 2000), found positive associations of small schools with grades, and the other studies also found positive associations with “academic performance” – these results are therefore consistent with the literature.
Retention

Student dropouts are a major problem facing urban school districts (Jimerson, 2001). Many students fail to complete high school (Rumberger, 1987), and this failure has serious long-term deleterious consequences on students’ futures (Roderick, 1994). SLCs might help alleviate this problem; for example, a greater connection to school and friends at school might encourage students to attend year after year, even in the face of poor academic performance (Cairns, Cairns, & Neckerman, 1989; V. E. Lee & Burkam, 2003). If SLCs help students stay in school even one more year, that provides another chance for a student to get help, earn credits, and eventually graduate (Roderick, 1994).

This analysis uses failure to proceed to the next grade as a proxy for dropping out because accurate dropout data was unavailable directly from the school district, and grade retention is an excellent proxy, according to Roderick, (1994). This failure can occur in two ways. Either students can stay in the same grade for more than one year (classic grade retention) (Shepard & Smith, 1989), or students can disappear from the data before they are in 12th grade as measured by credits earned (failure to graduate) (Jimerson, 2001). Given the available data for these two options, this analysis makes the assumption that students who disappear for any reason before the 12th grade are dropouts (instead of transferring or something positive), and that any students who make it to the end of 12th grade (as measured by their credits) graduated or will eventually graduate (instead of dropping out). These two assumptions are supported by three justifications: first, two alternate specification analyses (one alternately assuming only classic grade retention – held back or not; the other alternately assuming disappearance is “success”) yield similar qualitative results, while excluding many students, or stretching plausibility because as is common, (Jimerson, 2001; Shepard & Smith, 1989), over 80% of total grade retention across all
grades in this data happens in the 9th and 10th grades (but not necessarily the 1st and 2nd years of high school), and over 95% of students who earn enough credits to make it to the end of 11th grade appear at the end of 12th grade the next year; second, students who disappear are overwhelmingly from the lower end of the test score distribution, a additional risk factor according to Cairn (1989), making it far more plausible that, given the difficulty of moving entirely out of the large geographic area (over 700 square miles) of this school district, a student who disappeared had the negative final academic outcome, a result also found by Roderick (1994) and supported by Rumberger (2005); third, the first assumption aligns the best with the state’s definition of a dropout, which is: if their whereabouts are unknown, they are a dropout. Still, Table 6.4 shows the results of the analysis for the three different dropout specifications.

<table>
<thead>
<tr>
<th>SLC Implementation</th>
<th>Prob. of retention (disappear = dropout)</th>
<th>Fraction of students retained (disappear = dropout)</th>
<th>Prob. of retention (disappear = missing)</th>
<th>Fraction of students retained (disappear = missing)</th>
<th>Prob. of retention (disappear = success)</th>
<th>Fraction of students retained (disappear = success)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Base model)</td>
<td>-0.0110</td>
<td>-0.0226*</td>
<td>-0.00788</td>
<td>(0.00775)</td>
<td>(0.0101)</td>
<td>(0.00800)</td>
</tr>
<tr>
<td>(Full model)</td>
<td>-0.00812</td>
<td>n/a</td>
<td>-0.00787</td>
<td>(0.00732)</td>
<td>(0.00775)</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses. Significance: + = p < 0.10, * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

This analysis shows that the implementation of SLCs had a very small positive effect, regardless of specification, on dropouts/retention rates. While the individual measures of retention were not significant regardless of the assumed outcomes for those students who disappeared, the school level aggregate analyses yield decreases of between one and one-half to two percent in the number of students who are retained or who disappeared, depending on the definition. It is important to note for this analysis that the fraction of dropouts was dropping significantly throughout the district during this timeframe, and particularly in the school years of SLC implementation (Rayyes & Barela, 2009). Still, while not significant, the sign of the change in
individual probability of failing is negative, which corresponds well with the significant drop in fraction of students failing. These effects correspond well to the literature on small schools and SLCs, which also predict or find beneficial associations of small schools or SLCs with dropouts or graduation rates (Bernstein, et al., 2008; James-Burdumy, et al., 2008; Smerdon & Cohen, 2009; Wasley, et al., 2000).

Test scores
This chapter concludes with an analysis of student test scores. However problematic their interpretation or limitations, as a key component of current state and federal accountability systems their importance to school districts hardly needs stating. While SLCs do not always have any particularly new elements (such as new curricula, a different teaching method, or a renewed test preparatory focus) that claim to directly improve test scores above and beyond what school districts might be currently doing, SLCs might improve student learning directly and therefore test scores indirectly. For example, the professional development that is a desired attribute of SLC implementation, (Price, et al., 2003), could allow teachers to better coordinate their instruction, helping all students learn (Cotton, 2001). Improvements in the classroom climate (Cotton, 1996b), feeling connected (McNeely, et al., 2002), and the reduction in disciplinary problems found here and elsewhere, (Klonsky, 2002; Peterson & Skiba, 2001), could possibly allow teachers to spend more time on teaching and less time on classroom management (Wright, et al., 1997). A smaller set of students might even allow teachers to more easily modify instruction to help students who might otherwise fall through the cracks (Raywid, 1996).

The analysis of test score data is fraught with difficulty (Hamilton, Stecher, & Klein, 2002; Kane, Staiger, Grissmer, & Ladd, 2002). Due to previously stated methodological concerns
about the comparability of test scores given the testing procedures in this state, this test score
analysis conservatively focuses only on the English end-of-year test. To address the different test
score scales each year, the scores are converted to z-scores internal to the district (not the state)
as described previously in Chapter 1. This allows a more meaningful interpretation of test scores
as fractions of the grade level population standard deviation in test score.

This analysis uses four measures of test scores. The first two measures are indicators used for
state and federal accountability: whether a student is proficient or below basic, respectively. The
last two are direct analyses of the change in the per-student z-score and gain z-score due to the
implementation of SLCs. Results are tabulated in Table 6.5.

<table>
<thead>
<tr>
<th>SLC Implementation</th>
<th>Prob. Proficient +</th>
<th>Fraction of students proficient</th>
<th>Prob. below basic -</th>
<th>Fraction of students below basic</th>
<th>Change in z-score</th>
<th>Change in gain z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Base model)</td>
<td>-0.00223</td>
<td>-0.00449</td>
<td>0.00658</td>
<td>-0.000755</td>
<td>-0.0251*</td>
<td>-0.0102</td>
</tr>
<tr>
<td></td>
<td>(0.00525)</td>
<td>(0.00779)</td>
<td>(0.00556)</td>
<td>(0.00674)</td>
<td>(0.0123)</td>
<td>(0.0151)</td>
</tr>
<tr>
<td>(Full model)</td>
<td>-0.00169</td>
<td>n/a</td>
<td>0.00418</td>
<td>n/a</td>
<td>-0.0271*</td>
<td>-0.0161</td>
</tr>
<tr>
<td></td>
<td>(0.00497)</td>
<td>(n/a)</td>
<td>(0.00624)</td>
<td>(n/a)</td>
<td>(0.0144)</td>
<td>(0.0149)</td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses. Significance: + = p < 0.10, * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

This analysis shows the implementation of SLCs has no significant effect on English Language
Arts test scores. While all of the coefficients are negative, they only reach the level of
significance in one case which did not control for student-level factors. Even still, in that case,
the magnitude of the effect is very small, about two and one-half percent of a standard deviation
in test score. The analyses of the number of students scoring proficient and above or those below
basic indicate an almost imperceptible effect as well. Even considering the changes in retention
(and therefore composition), this is not practically significant. As perhaps might have been
expected theoretically, (Goldhaber & Brewer, 1997), and from the results from Chapter 5 and
other SLC research (James-Burdumy, et al., 2008; Shear, et al., 2008), however undesired, the
implementation of SLCs did not change students’ test scores or the amount students gained in test scores from year to year.

Section 6.4: Discussion

This chapter indicates that the implementation of SLCs had a small positive effect on students. The implementation of SLCs coincided with reduced suspensions, a smaller fraction of students failing almost all classes, and a reduction in the fraction of students retained in grade and “dropouts,” while having no significant effect on attendance. While these softer outcomes surrounding students that are more engaged with school do not receive as much attention as test scores, they are important outcomes for a good school and ultimately better citizens (Dewey, 1896; Hirschfield & Gasper, 2010). Disappointingly for certain advocates, as found by many studies of SLCs (Bernstein, et al., 2008; Shear, et al., 2008), the data from the test score analysis indicates that, at least for English, SLC implementation had an almost insignificant negative effect, only reaching statistical significance in one specification which did not control for student contributing factors. Other test subjects were not analyzed, and it is possible that the increase in numbers of students staying in school had an effect on this measure (e.g., if the students now staying in school due to SLCs’ better grade retention have relatively lower test score levels or gains compared to the student population). Still, the positive effects (keeping students in schools, with better grades and fewer discipline problems) of the implementation of SLCs in this district easily outweigh the null effect on test scores.
Conclusion

This research focused on school structure, composition, and one pathway through which the whole-school SLC reform could improve children’s educational outcomes: by changing schools’ and students’ course-taking structure so more students share more classes together, and reducing the effective size of the student cohort. This chapter provides brief answers to the research questions, followed by a more complete recap of results, policy recommendation, and future research directions.

Brief answers to research questions

RQ1. What is the pattern of performance across community detection algorithms?

   a. For high underlying community structure, all algorithms perform well.
      Algorithmic groups are very similar to known communities when known community structure is moderate or greater. The most complicated method, Reichardt’s (2006) Spin Glass model, was the best, but a “Fast Greedy” method by Clauset et. al. (2004) produced very good results and was much more efficient.

RQ2. To what degree are students in schools sorted or segregated?

   a. There was evidence of ability sorting, but little solid evidence of within school ethnic segregation. The amount of variation in academic outcomes within communities of students inside schools was equal to or up to two times greater than variation across schools.

RQ3. To what extent did the implementation of SLCs change school structure?
a. The implementation of SLCs significantly increased the amount of community structure and the number of classes students shared with each classmate, while decreasing the average number of unique classmates.

RQ4. How are measures of school structure related to academic outcomes?

a. Students who shared more classes with each classmate tended to have better academic outcomes by most measures. There was evidence that high achieving students formed coherent cohorts in upper grades, but had many unique classmates in the lower grades. There was evidence that higher achieving students took classes outside of their communities after SLC implementation.

RQ5. To what extent did the implementation of SLCs change academic outcomes?

a. The implementation of SLCs had modest positive effects, reducing suspension rates, the number of students failing most classes, and the number of students dropping out or retained in grade, while having no impact on attendance rates or test scores after controlling for student characteristics.

Methodology and composition

This research contributes a relatively novel way of viewing schools and students to the education literature. By viewing a high school as a network with students as nodes linked by shared course-taking, quantitative measures of structure and position are able to be calculated. Continuing further into methodology, this research rates five community detection algorithms, and concludes that for student course-taking affiliation networks of school size, unassisted community detection algorithms can successfully find high quality, relevant communities of students who take more classes together than with other students, a result found by a couple of other studies, but on a much smaller scale (Field, et al., 2006; Heck, et al., 2004).
The dissertation then addresses sociological aspects of high school structure related to SLCs. A network-theoretic measure of the level of association at the class level was presented, *assortativity*, and its results compared to a more traditional measure of segregation calculated across communities. Considering the composition and sorting question, this analysis finds that while individual schools differ in composition, on average, no evidence of ethnic grouping or segregation *within* schools, student communities, or classes exists in the district. But, while there is effectively no ethnic segregation, there is evidence that on average, modest ability-based grouping is present within schools, especially before the implementation of SLCs, as was suggested by Ready (2004), with potentially harmful outcomes for students (Braddock, 1980; Carbonaro, 2005; Chiu, et al., 2008; Frankenberg, Lee, & Orfield, 2003; Friedkin & Thomas, 1997; A. Gamoran, 1992; Adam Gamoran & Berends, 1987; Marsh, 1987; Marsh & Parker, 1984; Martin, et al., 2005; Russell & Gregory, 2005), and in fact, the amount of ability difference within school communities is larger that that which exists across schools, years, or grades, as also found by the Coleman report on racial segregation Coleman (1966) and other studies (Goldhaber & Brewer, 1997). In other words, the average difference between the highest and lowest performing communities in the average school is greater than the average difference between the highest and lowest performing schools.

**Small Learning Community effects and pathways**

The implementation of SLCs successfully changed school structure by modifying students’ schedules (Neubig, 2006), increasing community structure and the number of courses students share with each other, while also reducing the number of unique students each student must see in their classes each day.
Changes in the structural measures in the direction consistent with these structural effects of SLCs were correlated with improvements in most measures of academic performance, as also found by Weiss (2010), Ready (2004), and others. An increase in the average number of classes shared with each other student in classes is very highly correlated with positive academic outcomes (Bryk & Driscoll, 1988; Klonsky, 2002; Royal & Rossi, 1996). The total number of unique students a student sees across all classes tends to be positively correlated with academic performance in the lower grades, and negatively correlated in the upper grades – consistent with the hypothesis that higher level students branch out in lower grades, taking classes that their peers do not take (“passport” classes), and thereby being exposed to more unique students; when in the upper grades they tend to take more, smaller size classes with a smaller, coherent cohort than the average student, results supported by Delany (1991) and Oakes (1995) studies of traditional tracking. Finally, there is consistent evidence in the data that higher performing students were in groups taking classes together prior to SLC implementation as reported by Price (2003), and those groups are now indistinguishable from the school-wide SLCs – implying that prior research on selective schools, small schools, or SLCs, (Bernstein, et al., 2008; Cotton, 2001; Johns, 2008b; Raywid, 1996; Shear, et al., 2008), that failed to account for selection bias from unobservable characteristics (e.g., motivation), should be interpreted with caution.

Finally, although the precise cause cannot be definitively determined by this analysis design, an unfortunate but ubiquitous limitation in this type of research (Leithwood & Jantzi, 2009), on balance, the implementation of SLCs improved academic outcomes – having positive effects on discipline, GPA, and drop-outs – while having non-significant effects on test scores and
attendance rates in the full model specification. This dissertation corroborates non-causal research on SLCs and provides support for the idea that the small schools research might apply to SLCs.

Policy recommendations and future research directions

While the actual costs of this reform were not acquired, it is not a stretch to assume that this type of course structure change is inexpensive compared to building a new set of small schools (Dewees, 1999), which could also have these beneficial effects (Andrews, et al., 2002; Cotton, 1996a; Fox, 1981; Leithwood & Jantzi, 2009). The lack of an extensive research base, low relative cost and the presence of modest positive effects from this analysis inform the policy recommendation and future research directions:

Policy recommendation

- The Small Learning Community conversion reform should be continued, while continuing to look for other complementary methods to improve student learning.

Future research directions

- Similar research from a teacher’s perspective (i.e., teacher communities and shared conference periods) instead of the student’s perspective should be conducted (Febey, 2006; Wenger, 1998).
- Longer term research on academic outcomes and structural changes from SLCs should be pursued, including A-G (University of California and California State University entrance requirements) completion rates (Huizar, Lauritzen, & Tokofsky, 2005; Martin, et al., 2005).
- Research on the effects of SLCs on the multiple achievement gaps should be performed (Leithwood & Jantzi, 2009).
- Research on the effects of SLCs on the availability of higher level classes should be performed (Attewell, 2001).
Research on the effect of whole-school SLCs on the availability and access to arts, sports, and other special programs should be performed (Barker, 1985).

Research on which types of SLCs, including magnet schools and pilot schools, provide better outcomes should be performed (Kuo, 2010; Neild, 2004).

Research on best practices for developing an effective master schedule with SLCs should be performed and distributed to all administrators in charge of scheduling (Neubig, 2006; D. Ready, et al., 2004).

Experiments designed to determine which elements of SLCs are most effective should be performed (Price, et al., 2003).

Research correlating student feelings of connectedness with structural measures should be performed to further validate these structural measures (Calvó-Armengol, Patacchini, & Zenou, 2009; Gest, Farmer, Cairns, & Xie, 2003; Jackson & Rogers, 2007; Royal & Rossi, 1996).

The whole-school Small Learning Community structural reform comes tantalizingly close to getting something for nothing: if the benefits came purely from structural changes, theoretically it might only require low cost changes to schools’ master schedules and students’ schedules (Neubig, 2006). Determining which of the many other accoutrements of a “SLC” are required to gain the benefits found here is the key to a successful, efficient expansion of reform that could move education away from the brink, (Gardner, 1983), and help hundreds of thousands of urban public high school students in the United States of America.
References


Harms, W., & DePencier, I. (1996). Experiencing education: 100 years of learning at the University of Chicago Laboratory Schools.


Turnbo, B. J. (2007). *The relationship between small learning communities and student performance as identified by the Academic Excellence Indicator System at Robert E. Lee*


