

WP-98-24

**COMER'S SCHOOL DEVELOPMENT PROGRAM IN CHICAGO:
A THEORY-BASED EVALUATION**

Thomas D. Cook
H. David Hunt
Robert F. Murphy

Institute for Policy Research
Northwestern University

This research was made possible by grants from the MacArthur Foundation and the Rockefeller Foundation and support was provided by the Center for Advanced Study in the Behavioral Sciences, where the first author was a Fellow in 1997/98.

Abstract

Using 5th through 8th grade students, the Comer School Development Program was evaluated in 10 inner city Chicago schools over four years, contrasting them with 9 randomly selected no-treatment comparison schools. Comer schools implemented more program details than the controls but were not faithful to all program particulars. Student ratings of the school's social climate improved soon after the program began and, by the last two study years, both the students' and teachers' perceptions of the school's academic climate had also improved relative to the control schools. By these last years, Comer schools were also gaining about three percentile points more than the controls in both reading and math and students in them reported less acting out on a scale whose items are correlated with more serious offending in later life. Students in Comer schools also endorsed more conventional norms about misbehaving and reported greater ability to control their anger. However, the Comer program did not benefit either students' mental health or their participation in activities that adults consider wholesome. Explanations for the achievement and acting out results are offered based on student and staff data about school climate, on insights from an ethnography conducted in the program schools, and on contrast with the evaluation results from Prince George's County, Maryland, where a different variant of the program failed to achieve any positive outcomes.

INTRODUCTION

This is the second in a series of reports evaluating Comer's School Development Program (SDP) in different cities and suburbs. This report deals with Chicago and analyzes how well the program is implemented there, how it affects school climate, how it influences change in student outcomes, and why it has the effects it does. The first report was on the program in middle schools in Prince George's County, Maryland (Cook, Habib, Phillips, Settersten, Shagle & Degirmencioglu, 2000); and the next will be on SDP in elementary schools in Detroit (Millsap, Chase, Cook & Hunt, in preparation).

Program Description. James Comer contends that a wide range of student skills can be enhanced through an intervention of his design that seeks to improve the interpersonal relationships and social climate in a school as a prelude to enhancing its academic focus and hence student achievement. His theory has been outlined in detail elsewhere, both as it began and has subsequently evolved (Comer, 1980; Comer, 1984; Comer, 1988; Comer, Haynes, Joyner & Ben Avie, 1996; Anson, Cook, Habib, Grady, Haynes & Comer, 1991). Central to the theory is the notion that each school should determine its own academic and social goals. Specified are only the processes and structures needed to establish, monitor and modify these goals.

The premier structure is the School Planning and Management Team consisting of school administrators, teachers, other staff, parents and sometimes students. Its main purpose is developing a school improvement plan, eliciting support for this plan across the entire school community, monitoring progress and suggesting mid-course corrections. Also specified is a Social Support Team consisting of the school professionals most concerned with students' psychological and social welfare - counselors, nurses, social workers, special education teachers and psychologists. This team is supposed to minister to students with special needs and also to prevent social problems from arising in the building. It does all of this by disseminating to all in the building (and even to parents) what is currently known

about child development and its links to the social and racial attributes of local families. The final structure is the Parent Team. Its goal is to mobilize parents to support the school in various capacities - in governance, through service in classrooms, hallways and libraries, by supervising day trips and engaging in boosterism and fund raising. Comer hopes that Parent Teams will inject into schools the close community bonds that he thinks operated when nearly all teachers lived in the same community as their students.

These three structures are supposed to operate according to process principles that Comer trusts will become widespread throughout the building and not just restricted to the three teams. The first principle is that adult groups should cooperate with each other and always put student needs above their own. The assumption is that competing adults do not have the focus or energy required to serve children well. The second principle is that all adults in the building should adopt a problem solving rather than a fault-finding orientation. This avoids counterproductive acrimony and fosters improved teamwork. And finally, Comer highlights decision-making by consensus rather than vote or principal fiat. Consensus requires listening to others and developing empathy for others' views. It does not engender the polarization into winners and losers that voting can bring about. Once these three process principles are dominant in a school, Comer believes that the staff will then focus on attaining widely shared goals, interpersonal trust will be a premier building norm, and local children's needs will be well understood and obviously paramount. A more humane and effective school should result from all of this.

Comer distinguishes between physical, psychological, linguistic, cognitive, social and ethical pathways to healthy development. He believes that these pathways are interdependent and mutually reinforcing. This has several important implications. One is that promoting children's welfare in just one or two domains will rarely achieve anything of lasting value; and another is that evaluating his program requires testing children across many or all of these outcome domains. He also contends that teacher knowledge of students' backgrounds influences how they treat students, the implication being that the

social background of students will influence the kind of reform activities undertaken in a school and perhaps even staff's initial assumptions about their effectiveness. Most of Comer's own work has been with elementary schools serving students of color living in mid-size urban communities. In his view, these children face unique problems because their teachers are often middle class and of a different race. Because of this, as well as for other reasons already noted, Comer believes that special attention should be paid to improving a school's social climate. Achieving this should then reduce the teacher-student culture gap, help students acquire some of the middle class interaction habits that teachers (and employers) value, and make learning easier because students feel safer in school and trust the staff more. Because social climate is so central, Comer suggests using both the Social Service Teams and staff professional development time to help teachers better understand children's developmental needs and how these relate to local circumstances. He wants to reduce teachers' negative stereotyping of students, particularly those of color from poor households, and to get teachers to believe that their students can learn and that their parents are trying to support them under circumstances that often limit what they can do. Comer does not blame teachers for students' shortcomings. He also acknowledges that parents need to know more about the pressures school personnel face, need to take more responsibility for their children's behavior and need to do more so that children see how closely home and school values coincide.

Anson *et al* (1991) outline a causal model of Comer's theory that indicates what it takes to test the theory well. Program implementation is a first issue—describing and then evaluating how well the three teams operate, how widely governance is shared, to what extent the program's process goals are disseminated and by how much parent involvement is enhanced. Given quality implementation, the next question is: Does the school social climate improve, particularly as regards the quality of interpersonal relationships among staff, among students and between teachers and students? The next issue is to ascertain whether this improved social climate leads students to feel better about themselves and to behave in

ways that are closer to mainstream norms, including wanting to do well in academic work. Another issue follows from this: Does the school's academic climate improve when the social climate is better and children behave more conventionally? This then leads to a final question: Do higher test scores result from these climate and social behavior changes? Models like this tend to be neatly linear, though nature may be more complex. For instance, some of the postulated causal relationships may be reciprocal rather than uni-directional, and there is no specification of the length of time that has to elapse before a variable early in the sequence affects outcomes that are supposed to change later. Even so, the model points us to important domains where we have to look for effects, and it reminds us that evaluating presumed causal relationships later in the postulated sequence only makes sense if earlier ones have already been shown to occur.

The Policy Importance of the Program. The promise to bring about a broad set of positive developmental changes has led many foundations and school districts to support the School Development Program. According to Comer *et al* (1996), in 1990 there were 70 schools in the network. By 1995, there were 563 (433 elementary, 85 middle and 45 high schools) located in 80 districts and 22 states, including the District of Columbia. Many consequential claims have been made about the program's effectiveness. The 1998 Labor-HHS-Education Appropriations Bill included \$200 million for whole school reform, and the School Development Program was prominently mentioned as one of a very small number of reputedly already-successful school reform models. The program is also listed as "exemplary" in the U.S. Department of Education report: *Building Knowledge for a Nation of Learners* (1997). We suggest that the program is worthy of evaluation because it is long-standing, enjoys a national reputation, is growing rapidly, is especially relevant to minority students (particularly African-Americans) and because important claims have been made about its effectiveness in promoting healthy social and academic development.

However, nearly all the available data-based claims about program effectiveness comes from Comer's own research staff (summarized in Haynes, Emmons, Gebreyesus & Ben-Avie {1996}). Reviewing much the same evidence, Becker and Hedges (1996) noted the low technical quality of the research due to the way control groups were formed when they were available, to the biases associated with teacher self-reports of outcomes, to the high likelihood of capitalizing on chance in many of the statistical analyses conducted and to the small sample size of studies available for examining a given outcome. But within these limits, they suggested that the published results tended towards the positive, whether school climate, social behavior or academic achievement was under consideration. However, examining the same evidence and without splitting hairs, Cook et al (2000) preferred the verdict “not proven” over Becker & Hedges’ verdict of “possibly positive”.

To date, there has only been one independent study of program effectiveness based on a randomized experiment (Cook et al, 2000). It was conducted with 23 middle schools in Prince George's County, Maryland. The results suggested that the program was not implemented very well in this one county and had no effect on school climate or students’ mental health, social behavior or school performance. However, the study took place in a suburban county, albeit one with predominantly African-American schools into which many families have recently moved to flee the social disorganization of Washington, DC. Even so, the county’s poverty rates are not like Chicago’s. In the 1990 Census, no county tract had even 30% of its resident families poor whereas Chicago has been the site for many studies of urban underclass neighborhoods understood as having a poverty rate of 40% or more (Wilson, 1987; 1996). The Prince George’s County study also involved middle schoolers, though most of the national experience with Comer’s program has been at the elementary level where students are younger and typically attend the same school for longer. Also, Latino children had to be left out of the earlier study, even though the program is supposed to benefit all types of minorities, though African-Americans remain the major focus. A final limitation is that the earlier study had to use a state minimal competency test in math as the

academic achievement measure, though the particular test was developed to discriminate at the passing score rather than across the entire range. So, it had some scaling problems. Although the Prince George's County study involved schools formally in the School Development Program, the case could still be made that the student achievement measure was inadequate and that the sampling design did not include the types of urban schools and students for which the program was most clearly designed.

In the study to be reported, the sampled elementary schools were deliberately selected to be among the worst in Chicago. They serve mostly African-American students living in public housing, though some Mexican-American and Chinese-American students attend some of the sampled schools. The Iowa Test of Basic Skills was available in the city as an achievement outcome measure, and it is clearly superior for evaluation purposes to the Maryland state test. Also, individual student behavior and school performance were measured over four years rather than just the two years students attend county middle schools. The present study has a further advantage. Social programs are not static; they change in response to new personnel and experiences. For its first 25 or so years the Comer Program stressed improving school governance, dealing with problem children, enhancing social relationships in the building and raising parent involvement. Academic matters were secondary. That emphasis is now changing. Squires & Joyner (1996) provide a brief overview of new program thinking about stimulating academic achievement more directly. But even so, in a decentralized program it is still up to schools to select their own goals and procedures so long as Comer's process specifications are not violated. So, even with the newer priorities, Prince George's County officials could still have emphasized improving just their schools' social climate. Yet in Chicago as much emphasis came to be given to improving the academic focus as the social climate, thanks to exhortations from both the Chicago School District Central Office and program headquarters in New Haven. So, the present study gets closer to more recent program priorities.

The Program's Theoretical Importance. The Comer program is heavily psychological and developmental at its core. However, a heavy initial emphasis on social climate could have unintended negative consequences, especially if it takes time and effort away from more structured pedagogic activities (MacIver, Reuman & Main, 1995). For instance, many teachers learn about effective pedagogy during in-service training sessions. But if these sessions are devoted to learning about students' social development and mental health, to new ways of relating to parents, and to new teacher roles in school governance, then the in-service sessions are not likely to enhance academic achievement and may even detract from it. Moreover, if teachers spend more of their classroom time dealing with students' feelings, relationships and normative development, they may spend less time teaching the curriculum. In line with this, Phillips (1998) showed that efforts to improve a school's social climate can reduce academic achievement and Cook et al (2000) suggested the same in their Prince George's County study. Will this happen again in Chicago?

The issue is important because the notion of schools as learning communities is currently fashionable (Rutter, Maugham, Mortimore & Ouston, 1979; Bryk, Lee & Holland, 1993). In this approach, both the social and academic dimensions of school climate are emphasized because they are held to reinforce each other so long as the collective emphasis is clearly on learning. So, we will test whether a school's academic and social climate are improved because of the Comer program and whether these changes then mediate effects on individual students.

To summarize, we will describe and explain: (1) How well the School Development Program was implemented in Chicago; (2) Whether the program improved school climate and various student outcomes; and (3) Whether and how the social and academic dimensions of school climate mediate between program participation and outcome changes. Chicago is a good place for studying these issues. The schools in our sample are predominantly African-American and poor, students attend K through 8 schools, program

exposure lasts up to four years, the achievement measure is standard, and the local program eventually emphasized a school's academic focus as much as its social climate.

METHODS

The Setting and Program Variant. Elements of the implementation of the School Development Program are supposed to vary by both district and school in order to maximize the fit to local circumstances. The Chicago program is unique in many ways. It was introduced in the middle of a city-wide educational reform initiative supported by local politicians, businesses and philanthropies concerned about the poor performance of Chicago public school children. At first, the reform emphasized creating local school councils, usually composed of the principal and representatives of teachers, parents and community leaders (Bryk, Sebring, Kerbow, Rollow & Easton, 1998). It was assumed that these councils would govern schools better than the district central office or than principals acting alone because council members were held to be especially adept at identifying local needs and planning ameliorative initiatives that would be locally viable. Councils were encouraged to set their own school goals and establish their own instructional designs, thus entailing an overlap with the functions of Comer's School Planning and Management Team. But this problem did not arise in practice thanks to a division of labor that quickly evolved. Local councils retained control over budget and hiring matters, while the School Planning and Management Teams provided leadership in designing and monitoring other school priorities, including school improvement plans.

Midway through the study the political context changed. The mayor and his appointees in the Chicago district office tightened their control and reduced the power of local school councils. Some councils had been administered badly; others were co-opted by principals more anxious to retain their jobs than improve their schools; and others fell under the control of community activists seeking to use the schools as launching pads for broader reforms that were not educational. The new emphasis made test scores the dominant criterion for judging school effectiveness (rather than success in creating decentralized management), and one-sixth of all city schools were put on probation because of their low

test scores. The hope was that tighter management, a clear "bottom line", decisions linked to the bottom line and additional resources to each school would improve student test scores and so satisfy local taxpayers, encourage the State of Illinois to give more to city schools, and induce businesses and middle class families to stay in Chicago or locate there.

Comer was originally not keen to see his program introduced into Chicago, fearing it would be compromised by district, city and neighborhood politics. Nonetheless, he was persuaded to do so in 1990 by officials at Youth Guidance, a local organization of social workers that was then providing a broad range of non-instructional support services to high school students and their families. Expanding its mission, Youth Guidance eventually became responsible for implementing the Comer Program in Chicago K through 8 schools. The program was introduced in phases. In 1991-1992, four schools began as pilot institutions; next year, four more program schools were added (and four controls); and in 1993-1994, the final six program and control schools entered the study.

Youth Guidance's goals shifted over time. Much of the pilot and first full implementation year was spent stimulating parent involvement, getting principals and staff to see parents as important, and getting the three Comer teams to work in theoretically prescribed ways. Next year, greater saliency was accorded to improving social relationships, primarily between principals and teachers but also between various teacher groups in the schools and between parents and teachers. This interpersonal focus remained central, including in the second full year of implementation. But thereafter more and more attention was paid to improving academic performance through better instructional practice. To this end, Youth Guidance experimented for a year with a local university-based instructional program, but this did not work out. It then hired its own director of educational programming, but after initiating some preliminary efforts he was soon hired away by the School Development Program central office in Yale. It was only in the last two years considered here that systematic program efforts were introduced into schools to influence classroom change. These efforts involved changing some personnel, raising the profile of

instructional practice, modifying curriculum design and coordination, conducting more and better teacher training, and specifying the expanded roles that principals and master teachers could play in helping classroom teachers. These efforts increased the local correspondence with the shift towards instruction occurring in both the Comer central office and the city-wide reform initiative.

We do not have exact figures on program costs per school in Chicago. Most of these were borne by local foundations, though schools themselves came to contribute more and more out of the vastly increased discretionary budgets that were one consequence of the reform process. Nonetheless, program costs per school were clearly on the higher end of those found nationwide for the School Development Program. This was primarily because facilitators are a major share of any Comer budget and, in Chicago, they were originally assigned to a school full time--much more than the 1.5 to 2 days per week that occurred in, say, Prince George's County, and even more than is found in the many program sites where one facilitator looks after many schools (e.g., Cleveland and New Orleans). Interestingly, the facilitator expenses per school were not much higher in an absolute sense in Chicago than Prince George's County. This was because city facilitators were paid social work salaries, and these are less than half of what teacher facilitators received in the county. In-service training for teachers and parents is another large part of any Comer budget. The frequency of local retreats and of trips to Yale, the size of the contingents sent, and the use of educational consultants, all seem to us to be as heavy in Chicago as we have seen elsewhere in the School Development Program. On an annual per-school basis, the Chicago exemplar was probably among the better-funded nationally.

Thus, the Chicago variant of the School Development Program was unique in many ways. It depended on an agency staffed primarily by social workers and not trained educators. The facilitator role was particularly heavy. The School Planning and Management Team's functions overlapped considerably with those of local school councils. A very heavy emphasis was placed on parent mobilization initially and then on the quality of

social relationships in the school. And when it became obvious that a strategy was needed for directly improving achievement, many stops and starts occurred before such a strategy was in place. (To be fair, other Comer sites have never even adopted such a strategy, preferring to view the program only as a means of classroom management and social development.)

This local uniqueness does not imply a poor realization of Comer's theory in Chicago. After all, the theory does not specify who provides services and it frankly recognizes that program details have to be tailored to local exigencies. In fact, several additional factors point to a strong test of his model. One is the ability of Youth Guidance to learn and adapt Comer's model--as is attested by the School Development Program central office asking the agency to become a regional training center. Another is the level of financial resources available to each school in the program. A third is the dedication of Youth Guidance's senior staff, their continuity over six years and their skill in negotiating the perilous political waters of Chicago schools. A fourth is the willingness of the Chicago political, business and philanthropic elite to support school change. Another is the desire of many (but certainly not all) teachers and parents to embrace change. And the final reason is the ease with which the roles of the local school councils and the School Planning and Management Teams could be differentiated. All these factors point to the possibility of quality implementation of Comer's program in Chicago.

The Experimental Design Plan. The hope was to conduct a pilot year in four program schools and then, over two years, to initiate a multi-year randomized experiment with 10 matched pairs where one school from each pair was to be assigned to program status and the other to control. All the schools had principals who originally asked to be in the Comer program, and all were K through 8. However, only students in grades 5 through 8 were sampled for this study because of the need to complete written questionnaires.

The pilot sites were located in desperately poor African-American neighborhoods on the west side of Chicago. The principals wanted the program to help them restructure their school in ways that would be sensitive both to the city-wide reform mandate and to the racial and social class backgrounds of students. The eight Phase I schools were selected from a somewhat larger population of volunteer schools in the same area. Two years of prior data were used to match schools on the basis of average academic achievement and student racial composition before randomly assigning them to program or control status. A year later, the same procedures were used to assign a further 12 schools. These phase II schools were from somewhat less economically impacted parts of the city, largely because Youth Guidance wanted their facilitators to have some relief from the difficulties they were encountering working with the obviously troubled Phase I and pilot schools. Still, the Phase II schools were clearly urban, minority and poor. None had more than 2% White students, none scored above the 43rd percentile in math or reading on the Iowa Test of Basic Skills and, with the exception of one school, all the others had between 88% and 97% low income students. (The comparable figures for the Phase I schools were between 94% and 99%, omitting one outlier.)

In a meeting prior to any treatment assignment, all the principals agreed to abide by whatever assignment decision a coin toss produced and also to support the full measurement plan. They were also promised that if they became controls they could enter the program four years later if they so wanted and if financial resources also allowed. Although all the schools were majority African-American, some also had Latino (mostly Mexican-American) or Asian-American students (nearly all first- or second-generation immigrants from China). Schools with somewhat lower African-American representation tended to be in the second Phase.

Data were first collected from the Phase I schools in Fall 1992 and a year later for Phase II. In the first study year (or the second for pilot schools), questionnaires were administered to students in late Fall in order to provide a baseline for the written outcome

measures. Thereafter, outcome testing took place in the late Spring of each school year. About a month prior to this each year, students completed another questionnaire that dealt with school climate issues. School staff completed an annual questionnaire in late Spring that dealt with school climate and the quality of program implementation. This annual testing cycle continued until Spring 1997--five year after the study officially began for the pilot and first-phase schools and four years afterwards for the second-phase schools.

The Achieved Experimental Design. Unfortunately, five of the 24 schools dropped out of the study at different times for different reasons. One pilot school left the study just a year before its planned termination because Youth Guidance no longer wanted to work with a principal who had joined the school the year before and was not cooperating with them. Three of the original Phase I program schools also dropped out, each after two years. In two cases a new principal was appointed who did not want the program; and in the third, the principal did not want the quantitative research component (though his school stayed in the program and the ethnographic component continued). Of the Phase II schools, only one dropped out, almost immediately after learning it had been assigned to the control status. This left 19 schools--3 pilot program schools, 5 Phase I schools (1 program and 4 controls) and 11 Phase II schools (6 program and 5 controls).

Such selective attrition vitiates the randomized experiment since more Control than program schools dropped out and since the treatment is clearly confounded with principal turnover. We could restrict analyses to the planned school matches that were successfully maintained. But there is only one in Phase I and five in Phase II. Hence, we explored to see if any biases resulted from adding the pilot schools to the Phase I schools, thus creating 8 Phase I schools (4 program and 4 controls) and 11 Phase II schools (6 program and 5 controls). To do this requires treating the second year in pilot schools as the first year of legitimate data collection--a not unreasonable assumption since Youth Guidance spent much of the first year learning how to implement. Grouping the schools this way adds degrees of

freedom but entails the possibility of selection bias. To examine this possibility we first need to describe the student sample and the measures used for estimating possible bias.

The Longitudinal and Cross-Sectional Student Samples. With the school district's permission, all the original principals declared the research to be part of the school curriculum. So, only passive consent was required from parents. When a letter was sent to their home giving parents the chance not to participate, few parents chose to do so. Hence, the only student attrition was from those children absent on both the official testing day and the questionnaire makeup day, those who joined the school midway through the year, and those in the very small number of classes where a teacher refused to cooperate with the testing, despite repeated requests to do so by the school's front office. So, the student response rate was over 95% each year. Questionnaires were printed in English, in both English and Spanish, and in both English and Chinese (after back translation).

In total, 10,306 5th through 8th grade students were surveyed on at least one measurement occasion over the five year period. We refer to these students as the *cross-sectional sample*. Of these, 1685 were first eligible to be in the study in the fifth or sixth grade in the Fall of 1993 or 1994 and remained in the same school for three or four years between grades 5 and 8. These 1685 constitute the *longitudinal sample*. Of those with a four year history in the same school, 30% began in the first study year (1993) and 70% in 1994. Of those with a three-year history, 37% began in 1993 and 63% in 1994. These 1685 students constitute 68% of all those who began 5th or 6th grade in 1993 or 1994, the others being those who began then but left their school after less than three years.

This report uses the cross-sectional sample only to validate findings from the longitudinal sample. The hypothesis we test is whether program schools change over time at a faster rate than controls and so finish at a higher mean level. The longitudinal sample is superior for testing this because individual change is better estimated, annual fluctuations in

student body composition do not affect the school-level results¹, and the analysis includes the students exposed to the program or control status for the longest time possible.

Student Measures. The student measures are outlined in Table 1, together with some of their psychometric properties and a sample item from each construct. Many of the questionnaire measures are identical to those used in the evaluation of the School Development Program in Prince George’s County and some are being used in the ongoing evaluation in Detroit.

Insert Table 1 about here

1. Student Climate. A climate questionnaire was administered to students at the end of each study year. For ease of presentation, we organize the unique constructs into four domains. The first was Social Relationships with Adults in School. This consists of constructs assessing the respect with which school staff hold students and the caring students attribute to school staff.

The second was Social Climate Among Students. This consists of measures of students’ social skills, the use of positive problem solving strategies, the use of negative problem solving strategies, the perceived physical safety of students, pride in the school, attachment to the school, comfort in interpersonal contexts at school, and personal feelings of not belonging in the school.

The third was Academic Relationships with Adults in School. This consists of teacher’s encouragement of academic success, teachers’ concern with children’s learning, teachers’ behaviors to motivate students to learn, and the extent to which achievement is recognized in school assemblies.

The final domain sampled was Academic Values Among Students. This consists of devaluing academic performance, accepting school values and personally valuing education.

¹ Based on district records, the average student mobility rate in these schools is 30%.

Global social climate was assessed as the equally weighted sum of a sub-set of the constructs listed above: physical safety, respect of school staff for students, caring of school staff, students' social skills, the use of positive problem solving strategies, the use of negative problem solving strategies, pride in the school, attachment to the school, comfort in interpersonal contexts, and the prevalence of personal feelings of not belonging in the school. The global academic climate was computed by summing teachers' encouragement of academic success, teachers' concern with children's learning, teachers' behaviors to motivate students, recognition of achievement, student devaluing of academic success and student acceptance of school values.

2. Student Outcomes. Outcome constructs were also organized into four domains. The first was Mental Health. This consists of constructs assessing self-efficacy in school, racial pride, positive expectations for future life, valuing of positive outcomes in future, satisfaction with current life and relationships, the use of constructive coping strategies, depression, anger, and the use of negative coping strategies.

The second was Negative Social Behaviors. This consists of an "acting out" scale relevant to early teens based on asking about the frequency with which each of eleven mischievous or illegal acts had been committed in the last six months, items concerning normative beliefs about misbehavior, and questions about substance abuse, analyzed here as the sum of whether students had ever used cigarettes, tobacco or marijuana during the previous 30 days and had ever used more serious drugs during the last 6 months.

The third was Positive Social Behaviors, consisting of the proportion of all reported weekday and weekend hours spent in clubs or lessons, playing sports, doing homework or free reading.

The last was Academic Achievement, consisting of the ITBS Normal Curve Equivalents (NCE scores) for math and reading for every year a student had been in the Chicago school system. These data were provided by the school district.

3. Student Background. The student questionnaires included reports of the student's sex, race, parents' (or guardians') education, parents' (or guardians') current work status and household composition.

Also included were some family process variables, including measures of home academic support, use of reasoning in parenting practices, parental rule setting, and the quality of parent-child affective relations.

The combined 3rd grade ITBS scores for math and reading were also used as a student background variable since it predated the program.

School Characteristics. Any of the constructs above could be aggregated to the school level to characterize schools. But the Chicago Public School System also provided us with annual data on each school's enrollment, attendance, student mobility, percent of students classified as having limited English proficiency and the percent of low income students.

Sample Description. Table 2 compares the School Development Program and control schools on some official school demographics and test scores, plus selected student reports of family process and outcomes. The outcome data are from Fall of the first study year, and the student background and achievement are from Spring of the first year. The table reveals that the schools are very economically impacted and are almost exclusively minority. Most students have parents or guardians with no education beyond high school. More children live in homes without both biological parents than with them. Average test scores begin at about the 30th percentile, and their standard deviations suggest that very few students score above the 50th percentile. This is almost certainly the most economically disadvantaged large population with which the School Development Program has been tried.

Insert Table 2 about here

Selection Bias. Table 2 also reveals no initial differences between Comer and control schools in terms of school demographics and student reports of family demographics and process. The largest difference is in test scores where the controls initially outperform the program schools by about five percentile points in math and three in reading. There is therefore some selection bias favoring the controls, but it seems to be sporadic across the measures sampled. A similar conclusion is suggested by the school-level correlations between Comer status and school demography. In the school year beginning 1993, none of these correlations was significant across 19 schools, the highest value being .21, and the median -.05. Correlations with those outcomes that turn out to be key for the story to be told here are in the same range as the demographics: viz., -.20 for math, -.11 for reading, .02 for acting out, -.25 for normative beliefs about misbehavior and -.06 for lack of anger. These correlations suggest a selection problem, but of modest size. Further evidence of only modest bias comes from data on teachers. Across all years, the Comer schools average 55% Black teachers and 32% White, and the comparable control values are 57% and 29%. Of all Comer teachers, 17% were in their first year at the school, 12% were in their second year, and 31% had been there between 3 and 7 years. For controls these values were 15%, 11%, and 30%. Thus, the Comer and control schools were quite similar in student and staff demographics, though Comer students scored slightly below their controls on achievement tests.

More troubling is the potential selection bias arising from principal turnover. At the end of the 1993-1994 school year the Chicago School Board offered principals a generous retirement package and nine of the original 24 principals accepted. The next year two more did, one having spent only a year at his new school. The distribution of principal changes was correlated with program status. Of the ten Comer schools in the final analytic sample, only one changed principal during the study period while five of the nine control schools did. This differential was not a consequence of the program. After all, four principals eliminated the program when they began in a Comer school after the study was underway.

Hence, of the 14 planned program schools, five changed principals--the same number as in the controls. Even so, a potential selection threat is involved with the 19 schools we analyze if principal longevity (or an unobserved correlate of it) is correlated with factors causing school climate or student changes. For claims about program effects to be credible, the treatment-correlated difference in principal longevity has to be dealt with.

Staff Sample and Measures. On average, 968 staff members were surveyed annually (51 per school). Of these, 3% were administrators, 61% teachers, 11% non-classroom education specialists, and 25% supporting staff. 56% of the staff members reported being African-American, 26% European-American, 11% Hispanic-American, and 7% Asian-American. 67% of the teachers had been teaching in their schools for 3 or more years. Based on organizational charts provided by the schools at the beginning of each year, the response rate for teachers was estimated to be greater than 90%. The staff measures are outlined in Table 3, together with some of their psychometric properties and a sample item from each construct.

Insert Table 3 about here

Staff Ratings of Program Implementation. At the end of each study year, a school management and climate questionnaire was administered to school staff. A ten-item, theory-driven implementation index was used, the items from which are in Table 4. They assess the effectiveness of the School Planning and Management Team, the Social Service Team, the Parent Teacher Association, the school improvement plan and communication between teams. Also asked were items about the use of child development knowledge throughout the school, whether decisions were made by consensus, the commitment level of team members, the degree to which all members of the school community were included in decisions, and the extent of cultural and racial inclusion. Generic item wording was used because the control schools would not be familiar with vocabulary specific to the School Development

Program. Pilot work assured us that staff in Comer schools had no problems understanding this generic language. The 10 items were analyzed both as an index and separate items.

Insert Table 4 about here

Also included in the staff questionnaire were other implementation measures, asked because of their importance in Comer's theory. Prevalence of a No-Fault Orientation is one construct. Decentralized School Management is an equally weighted composite created from several highly correlated constructs at the school level: the extent of effort to implement decentralized management, the degree of success in implementing decentralized management, the extent of voluntary involvement in school governance, and the quality of the structures achieved for implementing decentralized management. Openness to Change is also a composite of several constructs: the extent to which communication channels are open, staff are open to innovative ideas, and staff are committed to change. Emphasis on Positive Multiculturalism is a single construct consisting of highly related items dealing with the quality of relations among staff of different racial backgrounds and the degree of multiculturalism in various school practices.

2. Staff Climate Ratings. The staff climate constructs were organized into the following domains: (1) School Social Climate, consisting of unique, multi-item constructs assessing teacher responsiveness to students, quality of relations within the school, teachers' pride in the school, non-teaching staffs' pride in the school, general attachment to the school, quality of school rules, interpersonal social skills of students, and the school-related behavior of students; (2) School Academic Climate, consisting of multi-item constructs about teachers' extra effort on behalf of students, teachers' responses to school problems around achievement, teachers' feelings of self-efficacy, the level of academic motivation attributed to students, and their expectations about the educational future for students; and (3) Parental Involvement, consisting of staff reports about parent involvement in general school activities, parent valuing of education, and interaction between parents and children about educational matters.

3. The Ethnographic Component. As part of this evaluation effort, Payne (1998) conducted an ethnographic study in each Comer school. Observations were made in the schools and at Youth Guidance office meetings, but interviews also took place. Visits were most frequent during the first four study years after the pilot schools began, and tapered off thereafter. In the earlier period, visits were about twice a week. Their main purposes were to describe (1) the operation of Comer teams and local school councils; (2) the organization and plans of Youth Guidance; (3) the behavior of Youth Guidance staff and facilitators at their headquarters and in schools; (4) the plans, behaviors and attitudes of principals and selected teachers, (5) the extent of parent participation; and (6) what happened in the retreats and in-service training that Youth Guidance organized for the school community at large or for various segments within it. The ethnographic component did not include systematic data collection about student behavior or in classrooms or in control schools. Nonetheless, it provides a record of what happened among adults in the program schools and in staff meetings about the program at Youth Guidance headquarters. It also serves to cross-validate findings about implementation and climate, to identify unplanned events and shifting program priorities, and to provide part of the explanation for any program effects.

Statistical Analyses

1. Students. For drawing inferences about Comer effects, the correct unit of analysis is schools. But since there are only 19 of them, statistical significance tests run the risk of false negatives. However, when individuals are the unit, the opposite risk pertains. So, to examine student outcome data we conducted both school- and individual-level tests. In our judgment, no simpler analysis plan is feasible for sailing between the Charybdis of inappropriate stringency and the Scylla of irresponsible liberality. We considered using the data from all 10,306 students in a single hierarchical linear model (Bryk, Raudenbush & Congdon, 1996). But we were leery of this since, by design, almost half of the students provided data at only one time point and these students were over represented at the first and

last assessments that are so crucial for estimating treatment effects. So, we place most confidence in results from school-level analyses of the longitudinal sample that estimate reliable slope or mean differences between the treatment and control schools. We place less confidence in slope or mean coefficients that are only reliable in individual-level analyses.

To protect--at least in part--against selection differences due to school attrition, the analytical model includes many individual difference measures - race (African-American versus Latino, Asian-American, and Other [mostly European-American]), gender, family composition, employment status of parents or guardians, and third grade combined ITBS test scores in math and reading. Since Phase II schools were deliberately selected because they are located in less depressed neighborhoods, we also use phase as a binary school-level covariate, treating the pilot schools as Phase I institutions. Selection is not likely to be a large problem in this study because of the partially implemented random assignment, the small differences in Table 2, the extensive statistical controls at the individual level, the formulation of hypotheses as treatment-related differences in change, and the chance to compare results with Prince George's County where random assignment was better preserved. However, treatment-related principal change remains a problem.

We tested hierarchical linear models that use occasions of measurement within individuals as the first level, individuals within schools as the second, and schools as the third. For level one, the last observation (8th grade) was chosen as the zero reference point so as to estimate mean differences in the last study year. The three-year longitudinal students are inevitably missing 5th grade observations or 8th grade measures, and these were estimated using the procedures built into the HLM program. Gender, race, household composition, parent work status and third grade achievement were used as covariates at the second (individual) level and any missing data were estimated using EM algorithms. (There was a severe missing data problem only for third grade achievement--just under 30%.) The school-level covariates were Comer, Phase and the Comer X Phase interaction. In

estimating differences in slope and mean, the individual and school models were identical except for whether school effects were specified at the individual or school level.

This analytic model is stronger for examining treatment differences in student outcomes than school climate. Since climate was measured at the end of each school year there is no Fall baseline measurement that occurred very shortly after the program was introduced and before it is plausible to assume program effects. The absence of a baseline creates obvious difficulties. For instance, if there is a first-year program effect on school climate, then the treatment and control means will differ at the first testing. But without convincing random assignment we could not be sure that the difference is due to selection rather than the program. Although we will model selection using an unusually long string of plausible student variables and one school-level control variable, there is no way to know that such a model is perfect. Inferentially, the situation is somewhat better if the treatments differ in climate slopes. But even then, selection is not completely ruled out, for the two populations may vary in their trajectories in ways that are not captured by the temporal assumptions built into the covariance-type analyses we conduct. So, the most interpretable pattern of climate results would be if the Comer and comparison means fail to differ in the first post-treatment years but come to differ thereafter.

2.Staff. The staff implementation and climate data could not be analyzed like the student climate data because, at the principals' request, no individual staff identifiers could be attached to staff surveys. To model effects at the school-level, two-level HLM analyses were conducted that use repeated staff cross sections in each school. So, the first analytic level involved staff responses aggregated to the school level for each of four years; the second level was that of the school where the predictors were Comer, Phase and Comer X Phase. A program effect is suggested if program schools change at a faster rate than controls and if, by the final time point, program school staffs view climate and implementation as more positive than control school staffs.

To estimate staff effects at the individual staff level, OLS analyses were conducted. These tested a model with two staff control variables--staff race (Black, White, Latino or Asian) and role (administrator, teacher or support staff)--and the usual Comer, Phase, Comer X Phase effects. Also included were the interactions of year with each individual difference and school factor. This is a reasonable approximation to the HLM model, but differs in one respect other than the validity of the standard errors. The interactions with year were highly correlated with the main effects constituting them and so year had to be centered at the midpoint of the time scale rather than the end. To equate the time at which a staff treatment effect is inferred, one needs to add to the mean treatment effect estimated from the individual-level analyses one and one-half times the estimated slope difference.

RESULTS

How Well was the School Development Program Implemented? Implementation was assessed three ways. One was on the 10-item scale that is weighted towards staff judgments about teams, shared decision-making, school improvement plans, human development principles and racial inclusion practices. Across four years, the scale alpha averaged .89. The measure assumes that (1) teams have already been set up and meet regularly--a common and gratifying structural finding in studies of the School Development Program (Cook et al, 2000; Millsap et al, 1998); and (2) the program's major purpose is to improve the school's social rather than academic climate--the clear program priority when this evaluation started; and (3) the quality of the facilitator is not crucial in differentiating between Comer and control schools because the major part of the facilitator's role is to create better teams, to enhance collaborative decision-making, to teach and exemplify what better child development and mature interpersonal relations are and to mediate between the different teacher race groups in the school. In any event, an implementation measure like the one above has now been used in Prince George's County and Detroit and, in a slightly

modified version, at Yale where the School Development Program research staff is using it in its own implementation studies.

Figure 1 graphs the four years of results for the index. Both the program and control schools start at about 2.9 on a 5-point scale, indicating that they were both somewhat Comer-like even before the program was fully introduced. Both time trends then increase over time, suggesting that all the Chicago schools were becoming more program-like, whatever their formal treatment assignment. The program slope is slightly steeper, though, with program schools initially outperforming controls by about .04 units and subsequently outperforming them by about .12 units. However, the first row of Table 4 shows that none of the treatment-related mean or slope differences were reliable. Striking is the similarity between Figure 1 and the same figure from Prince George's County (Cook et al, 2000). At each site, the program and control schools start at about the scale midpoint, each group increases over time, the increase is steeper in the program group--but not reliably so, and the final Comer mean is far from the scale endpoint.

Insert Figure 1 about here

Although there were high school-level correlations among the items constituting the implementation index, it is still possible that the index obscures a more differentiated pattern of effects at the item level. Table 4 reports the results of statistical analyses at this level, and Table 5 shows the corresponding means. The tables reveal that decisions are more often made by consensus in Comer schools, that decision-making processes there may more often involve the entire school community, and that between-team communication may be better there. Since these are the only reliable effects at the school or individual level, the average Comer school is not rated as having more effective teams or using human development principles more frequently or having a more effective school improvement plan or being more racially inclusive. (These results are based on the entire school staff but replicate in analyses limited to just teachers and administrators).

Insert Table 5 about here

The preceding results depend on single items of inevitably questionable reliability. In contrast, the bottom part of Tables 4 and 5 gives results from multi-item scales. These show improvements over time in both program and control schools in the quality of decentralized management, the prevalence of a no-fault orientation, the school's openness to change, its multiculturalism and its level of parent involvement. For some of them, though, the rate of improvement was steeper in Comer schools. Specifically, teachers in Comer schools reported that they spent more time interacting with parents about their children, though they did not see parents as valuing education more or as being more involved in volunteer activities. Comer staff also reported that more efforts were made in their schools to achieve decentralization, that staff members were more involved in governance activities and that greater success was realized in achieving decentralization. However, the School Planning and Management Team was not seen as instrumental in bringing this about; nor was there evidence that a no-fault orientation was more prevalent or that Comer schools were more open to change or that multiculturalism was more developed

Evidence on implementation also comes from the ethnography (Payne, 1998). It concluded that many of the planned implementation events actually occurred. In almost all program schools the three teams were set up; the local school council and the School Planning and Management Team were smoothly demarcated. Parents participated in more social events than they had done earlier (especially in the first study years), and they also participated somewhat more in governance and daily school activities. The Social Service Teams operated well for individual student referrals, though less well for fostering a school-wide prevention orientation. And in the program's first years, little emphasis was placed on instructional reform, though after some initial exploration new teaching practices were eventually implemented concerning teacher training, classroom teaching and principal support for improved instruction. Also very salient in the ethnographic account is the variation in implementation quality, something the standard errors in Table 4 also suggest since they are nearly 10 times larger at the school than the individual level. Some of this

heterogeneity is due to differences between Phase I and Phase II schools, but considerable variation is also found within phases. Unfortunately, by the end of 1997 the ethnographers were not willing to classify any school as faithfully following all the program guidelines, though some were rated as close.

Implementation Summary: Taking the quantitative and qualitative implementation findings together indicates (1) that Chicago was in the midst of a system-wide reform effort that may have fostered the implementation of Comer-like activities in many city schools, including the control schools that originally volunteered to be in Comer's program but lost out in the coin toss; (2) that some program elements were implemented better in Comer schools, particularly setting up the three Comer teams, getting more parents involved with the school, getting more school constituencies involved in decision-making and deciding more often by consensus rather than by vote or principal fiat; (3) that other major program elements were not implemented better in the average Comer school, especially as regards the quality of the school improvement plan, the quality of team functioning, the use of human development principles, openness to change and improving race relations among the staff; (4) that the Comer schools varied considerably in how well they implemented program details; and (5) by late 1997, implementation was not perfect in any program school. These results are similar to those from Prince George's County, and they have two major implications. The first is that, while the School Development Program changes what happens in schools, it does not transform them; and the second is that, it will be difficult to demonstrate program effects within the current experimental framework if implementation quality mediates student change. This is because such quality does not vary much by Comer status. However, if other causal processes are at play that are correlated with Comer status--like changes in social or academic climate--then program effects could still result. In this connection it is worth remembering that facilitators worked directly to improve the social

climate of the school, and later on its academic climate. Not all of their efforts were directed at promoting the teams and principles on which Comer's program depends.

Effects on School Climate.

1.Social Climate. School climate was measured on both staff and students. The statistical results from the staff climate measures are in Table 6 and the relevant means in Table 7. Salient are the many positive temporal changes, indicating a steady improvement across nearly all the school climate areas studied. However, in neither school- nor individual-level analyses did the coefficients indicate a steeper rate of change in the Comer schools. So, Comer staffs saw their climates as improving, but at the same rate as in the control schools. However, many final mean differences were reliable, though only at the individual level. Their direction suggests that Comer staff finished up seeing their social climates as less desirable.

Insert Tables 6 and 7 about here

This can be seen most clearly by averaging all ten social climate ratings for each school --each on a 5-point scale--and then computing Comer and control means. Figure 2 results. Evident is a positive change in social climate in all schools, but the mean is consistently higher in the average control school. However, we have to caution against over-interpretation here since this mean difference is not reliable at the school level, the random assignment was not perfect, there were no baseline social climate measures and the statistical analysis has only two individual difference controls--for a staff member's race and role.

Insert Figures 2 and 3 about here

Fortunately, the story is clearer for student social climate. The analysis results are in Table 8 and the means collapsed across grade are in Table 9.² By the criteria outlined earlier, final mean differences can be inferred for: attachment to school, valuing of school,

² OLS analyses using four grade levels as well as phase and Comer status showed some interactions of Comer and grade. But they were hardly more than would be expected by chance, and they did not replicate within domains. Hence, we proceed assuming that the treatment effects we discuss are not grade-contingent and so summing across grades is warranted.

valuing of education, the caring attributed to teachers and the social skills attributed to students. When all 10 student social climate constructs are averaged (and not just the reliable ones noted above) Figure 3 results. HLM analyses support three findings: (1) as students in the longitudinal sample age, they tend to see the social climate as deteriorating; (2) social climate is consistently higher in the Comer schools, with the mean difference being marginally reliable in school-level analyses and reliable in individual-level ones; and (3) no reliable slope differences emerge in any analysis. Thus, a first-year effect may have occurred for student social climate that persisted over four years.

Insert Tables 8 and 9 about here

2. Academic Climate. The first three measures from the staff academic climate domain assess teacher reports of their behavior and of their beliefs about classroom-based practices likely to stimulate achievement. Each shows that the teaching climate is improving across all schools; that the slope differences favor the Comer program; but that none of these differences are reliable. However, summing the three constructs results in Figure 4, with its small immediate advantage for the control schools and a steeper rate of change in the Comer schools so that they eventually catch up with the controls. The slope difference is now reliable, even at the more stringent school level of analysis. The other dimensions of academic climate, dealing with staff expectations about students, do not reveal the same trends. This suggests that the program affected teachers' views of their own behavior likely to stimulate achievement, but it did not affect their views of students' academic capacities and motivation.

Insert Figures 4 and 5 about here

As for student ratings of academic climate, the program and control means were not different in the first (or even second) study year but came to diverge thereafter. In both school and individual-level analyses, program students came to see academic achievement as increasingly more honored in school life; they came to see their teachers as ever more concerned with learning and as trying ever harder to motivate students academically. In

addition, they came to see their classmates as increasingly more accepting of school values and (perhaps) as devaluing school less.

To facilitate generalization, we created a student academic climate composite by averaging six of the seven student academic climate measures, whether they had a reliable slope difference or not.³ Figure 5 resulted. It showed that, as students in the longitudinal sample age, they see the academic climate as deteriorating--but more so in the control than the Comer schools. Consonant with this, HLM analyses showed both slope and final mean differences that were only marginally significant at the school level but were conventionally reliable at the individual level. Thus, from the student perspective, Comer's program seems to have raised the profile of academic issues and made students more disposed to achieve. However, as with the staff academic climate measure, it took several years for this awareness to grow.

3. Climate Summary. The academic climate results are similar for both teacher and student reports. In each case, a marginally steeper rate of change occurs in the Comer schools, whether they start at the same point as the controls (student reports) or at a slightly lower level (teacher reports). This replication occurs even though the academic climate constructs differ by respondent. Teachers respond to questions about what they do in the classroom to stimulate students, whereas students respond to items about academic relationships among students and about academic relationships with staff that are not limited to the classroom.

The social climate results are less clear. In the student data, Comer schools do better than controls by the very first post-treatment assessment and maintain this advantage thereafter; but in the staff data, the control schools tend to outperform the Comer schools both initially and subsequently. Interpretation of the student results as an early onset effect is potentially clouded because in these analyses climate was assessed at the end of the third year for pilot schools and the second year for the true Phase I schools. Only in Phase II schools was the first measure at the end of the first program year. So, the same analyses

³ Personal Valuing of Education was left out because of its conceptual redundancy with Student Acceptance of School Values.

were conducted with just the Phase II schools, and they resulted in the same data patterns. Although these results suggest that students in Comer schools noted an immediate and persisting change in their school's social climate, there are several reasons why we need to be cautious about this conclusion. The analyses did not include true pretests, the selection controls were not perfect, the same effect was not observed in Prince George's County or with Chicago staff, and there were no treatment differences in slope, only in means. However, we can be more sure of the positive academic climate effect at the later time period, given how initially similar the Comer and control schools were and the fact that both the student and teacher data showed steeper rates of increase in Comer schools.

Effects of the School Development Program on Student Outcomes. Table 10 presents the relevant analysis results and Table 11 the means. We begin with achievement test scores. As depicted in Figures 6 and 7 for math and reading respectively, the unadjusted means for the longitudinal design show that, for both reading and math, program schools start about 3 points below the controls and finish up level with them, for a net benefit of 3 points.⁴ After HLM adjustments to take (partial) account of the initial differences, the longitudinal design shows a gain of 1.39 units per year in math, equivalent to 4.2 points over three annual change periods, and of .99 units in reading, equivalent to about 3 units. Taking the error of these estimates into account entails a total gain of between 1.7 and 6.7 units in math and between 1.1 and 4.8 units in reading.

Insert Tables 10 and 11 and Figure 6 and 7 about here

By conventional standards in psychology (Cohen, 1988), these mean changes are "small" relative to their standard errors. But these are among Chicago's poorest, minority schools, any gain is probably a real accomplishment here given how recalcitrant such schools have proven to change attempts over the years. Moreover, there is active dispute

⁴ The size of the initial achievement difference is slightly different in Tables 2 and 11 because the population in Table 2 is all those students providing data at the beginning of the school year, whether in grade 5 or 6. Table 11's population is all 5th graders providing data.

about whether such gains should be labeled “small”. Rosenthal and Rubin (1994) have documented how so-called small changes can result in large effects when translated into different metrics or summed across individuals, while the changes observed in the famous Tennessee class size experiment that led to state-level policy changes are smaller than this in effect size terms (see Finn & Achilles, 1990; Mosteller, Light & Sachs, 1996) where the effect is .17 for the sample of classrooms at large and Ferguson (1998) where it is about .25 for African-American students only. It is also worth remembering that we obtained an annual slope difference over only four years. Were this one point NCE gain to be maintained over each of the eight elementary school years and then across all four high school years, the total effect would be of such inter-ocular significance that no discussion would be needed of the meaning of the size of the achievement results.

The next outcome domain we consider is negative social behaviors. Students in Comer schools reported relatively less acting out over time. The relevant index, which includes many items that correlate with criminal behavior in late adolescence and early adulthood, measures the frequency of performing 11 mischievous or delinquent acts in a variety of school and non-school settings. As is clear from Figure 8, all schools report more acting out as students age. But the rate of increase is less steep in the Comer schools. Examining the coefficients for slope differences shows a change of .19 units per year, equivalent to .57 units over three gain periods. Since the actual difference at the last data point is of .45 units, this suggests that the Comer group started a trivial .12 units below the controls and that, by the study’s last year, program students were committing about half a bad behavior less per year. The acting out index was also decomposed to assess acting out in school (3 items), committing mischievous acts outside of school (4 items), and breaking the law (also 4 items). Trends were in the same direction for each type of acting out, though the statistical results were clearer for delinquency and mischief out of school than for mischief in it.

Insert Figures 8, 9 and 10 about here

Three other findings buttress the conclusion that the Comer program caused a general decrease in negative behaviors. One is the change found in mainstream beliefs expressing disapproval of misbehavior. The relevant tables and Figure 9 show that students in Comer schools came to adopt ever more conventional beliefs about misbehavior when compared to controls. Another is the means for using illegal substances--drugs, tobacco and alcohol combined. These are also congruent with a positive program impact, though the effect is very small and only reliable in individual-level analyses. And finally, the program affected a measure of the frequency of getting angry across four situations that may well be conceptually linked to acting out. As Figure 10 makes clear, and as both individual and school-level analyses attest, students in the Comer schools came to report anger less frequently than controls, especially in the last study years. Thus, the weight of the evidence suggests that the Comer program reduced negative social behaviors and also modified some beliefs and anger feelings that might be causally related to acting out. These effects were clearer in the last two study years. Since the Comer and control schools did not differ before that, it is easier to defend the claim that selection probably does not account for the effects on self-reported negative behaviors.

Although the program reduced reports of such behaviors at a faster rate in the Comer schools, there was no corresponding increase in positive behaviors. This was true whether the measure was attending organized after-school activities in schools or in the neighborhood or spending more time in solitary activities that adults typically consider to be wholesome--like doing homework or reading. Moreover, the program also does not seem to have had any positive effects on mental health other than dealing with anger. Thus, the program seems to have reduced negative behaviors without enhancing positive ones or improving mental health.

The foregoing results do not take account of the possible selection confound due to Comer schools having more stable principals. Table 12 shows what happens to achievement gains, acting out and conventional beliefs about misbehavior when the schools with stable

principals are partitioned by Comer and control status. Since the program schools outperform control ones even in schools with stable principals, it seems unlikely that group difference in principal stability can alone account for the obtained effects.

Insert Table 12 about here

Another potential selection confound follows because in the penultimate study year the district central office announced that it would put schools on probation if 85% or more of their students scored below national norms. If no subsequent improvement occurred, schools could then be closed or reconstituted, threatening teachers' jobs. Six of the 10 Comer schools went onto probation versus only three of the nine controls. Although the probationary schools initially under performed the others--by about 6 NCE points in math and about .35 acting out behaviors, outcome changes were no larger in the Comer schools on probation than they were either in the non-probationary Comer schools or in the control schools on probation. Thus, treatment-differences in probation are not a plausible alternative interpretation of the current findings. All the schools were so low-performing that narrowly avoiding probation in one year did not absolve any school from the immediate need to improve test scores lest probation be their fate next year.

Improving achievement and decreasing acting out are outcomes of great significance in current national debates about effective schools. So, we need to ask whether the schools showing achievement gains were the same ones preventing negative social behaviors. Using the longitudinal data from just the Comer schools, Table 13 presents the school-level correlations between linear changes in the various outcomes showing program effects. Changes in acting out, normative beliefs about misbehavior and lack of anger are strongly related to each other, as are changes in math and reading. But the acting out and achievement domains tend to be negatively related. Within limits imposed by having so few schools, the data suggest that the schools where academic achievement changed more were not necessarily the same schools that successively reduced negative social behaviors. The story is similar at the student level. The students attending Comer schools who were changing

more in math and reading were not those with the most positive changes in acting out and related beliefs. So, it is not easy to argue that individual Comer schools or students benefited from a “double whammy”--both higher achievement and less acting out.

Insert Table 13 about here

Replication with the Cross-Sectional Sample. The cross-sectional design involves repeated cross sections of 5th through 8th graders in each school for the last four years of the study. The design is based on four years of data, four grade levels, 19 schools, 10,306 independent students and up to 16,952 observations per outcome. However, the cross-sectional sample includes students who, for whatever reasons, were in the same elementary school for only one or two years during our data collection. Thus, 49.5% of students provided data at only one time; 27.5% twice; 16% at 3 times and 6.9% at four. The number of repeat observations depends on three major factors. The most important is a student’s grade when the study began—e.g., an 8th grader in the study’s first year and a 5th grader in its last year can only provide data once. Another is the rate of turnover out of a school. And the third is new entrants into a school after 5th grade who could not have experienced any program exposure in earlier grades. Indeed, by the last study year students in the longitudinal design had been in Comer schools for 4.39 years on the average (remember Phase I students could have been in them for five years), while cross-sectional students had been in them for only 3.19 years. So, if length of program exposure mediates program effects, the cross-sectional design is less powerful than the longitudinal one. However, exposure is not the only theoretically relevant mediating process. If the Comer program generated some emergent school property that influenced even new entrants--say, a new norm about the importance of school or of academic achievement--then this norm change might influence any newcomers without diluting the program effect. While both the longitudinal and cross-sectional designs test for treatment differences in slopes and means, this does not imply identical populations or mediating processes.

For school-level cross-sectional analyses a three level HLM model was used. Student responses were aggregated to the grade level by school and year, thus creating 304 grade/school/year units (4 grades by 19 schools by 4 years). At the first level, the model represented multiple measurement occasions (years) nested within grade/school units, and the last study year was chosen as the zero reference point. At the second level were the average 3rd grade achievement means for each of the 76 aggregated school/grade combinations. Other potential covariates at this level were not used because, within schools, there was little annual between-grade variation in average gender, race, household composition or parent work status. At the third (school) level, the predictors were Comer, Phase and the centered Comer X Phase interaction.⁵ Thus, the analysis probes treatment differences in final means and four-year slopes across the 5th through 8th grades in the average Comer and comparison school. To model student effects at the individual level, the cross-sectional design used ordinary least squares (OLS). The predictors were gender, race, grade, phase, year, prior achievement, family composition, parental work status, Comer status, Comer X Phase, plus the interaction of year with each individual difference factor as well as with Comer, Phase, and Comer X Phase.

Some of the cross-sectional and longitudinal results cross-validate. Figures 11 and 12 present the cross-sectional findings for social and academic climate. By the end of the first year, there is a positive program effect for social climate that persists over time. The Comer and control schools start out equal for academic climate but a reliable mean difference emerges by the last year that favors Comer schools. Both of these climate results replicate what was found longitudinally.

Insert Figures 11 and 12 about here

Figures 13 through 15 present the results for the social behavior variables. The means for acting out, disapproval of misbehavior and anger control suggest the same pro-Comer finding as in the longitudinal design, though now none of the differences are reliable

⁵ The analysis takes no account of grade differences in the size of any Comer effect because we checked for this at an early stage in the analysis and could find no more than would be expected by chance.

at either the school or individual level. Unfortunately, neither the cross-sectional means nor analysis results indicate any systematic differences between the Comer and control schools for math or reading (see Figures 16 and 17). So, the climate results replicate better across designs than do the outcome results.

Insert Figures 13 through 17 about here

The design difference in achievement results requires that students uniquely in the cross-sectional sample manifested Comer/control time trends opposite to those obtained with the longitudinal sample. Indeed, students who were in Comer schools for just a few years tended to score lower than their counterparts in control schools; those leaving Comer school after a few years tended to score higher; and younger grades entering the sampling framework in later years tended to be less able if they were in the Comer than control schools. Consequently, Comer students not in the longitudinal sample decreased in achievement relative to their controls while the opposite occurred with those in the longitudinal sample. We do not know why this was the case, but it must lower somewhat our confidence in the longitudinal achievement results presented earlier. It does not reduce the confidence to zero, however, given the widely acknowledged superiority of longitudinal designs for ascertaining program effects because units are exposed to the independent variable for longer and because there are no treatment-related perturbations associated with time-varying population shifts.

The Mediating Role of School Climate. When the average Comer and control schools are compared, the longitudinal results are congruent with Comer's theory. That is, student social climate is immediately superior in the Comer schools (Figure 3); in both student and staff accounts, academic climate subsequently improves (Figures 2 and 3); and some student outcomes change for the good, but again mostly at later time points (Figure 4). The data seem to mimic Comer's theory almost perfectly. Hence, if explanation depended only

on contrasting the average Comer and control schools in a one degree of freedom test, the task would be easy and the results clear-cut.

However, individual schools can be compared and not just averages of schools. So, we used the student social and academic climate indices in what was otherwise the same longitudinal model that generated the school-level Comer effects in Figure 4. The aim was to see if the slope differences favoring Comer disappeared once the climate variables were separately introduced into the HLM analyses. The school level predictors were Comer status, Phase, the Comer X Phase interaction and, depending on the analysis, another predictor was added. It was either the school's social or academic climate, this climate was expressed either as a mean or a slope, and it was assessed either from student or staff climate reports aggregated to the school level. When slopes were in the analysis, the initial starting mean was also added. Unfortunately, the treatment differences in academic achievement and acting out slopes did not change much when any of the climate variables were introduced into the analysis, suggesting that the Comer schools where individuals changed more were not necessarily those with better social or academic climates⁶.

Comer's theory suggests that social climate should improve prior to a more positive academic climate emerging. So, longitudinal models were constructed that simultaneously introduced into the usual outcome model both the mean social climate across all four years (either the student or staff version) and the change in academic climate during the last two study years (again, either in the student or teacher versions). The aim was to describe a setting with a good social climate and an improving academic one. But such models also failed to make the effects in Figure 4 go away, though one of them did marginally reduce the coefficients for acting out and normative beliefs about misbehavior. In general, though, the pattern of outcome changes obtained when contrasting the average Comer and control school does not mimic how individual Comer schools change. Schools with more favorably

⁶ We did not use the instrumental variable approach to mediational analyses suggested by Angrist, Imbens & Rubin (1996) since it relies on a successfully implemented randomized experiment. While our was largely achieved, it was not perfectly achieved.

changing outcomes were not those with consistently better social or academic climates or with better combinations of climate components. Other factors are likely responsible for the pattern of obtained program effects.

In this connection, the school-level correlations among the climate indices are instructive. For the Comer schools, the mean student social and academic climates are correlated .79 and the two slopes .68. (The corresponding values for staff are .43 and .60.) These high values indicate how difficult it is to discriminate between the social and academic aspects of school climate on empirical grounds alone. The two are similar, though as the different relationships with the Comer manipulation illustrated, they are not identical. Nonetheless, it will be difficult to uncover independent social and academic climate effects. It will also be difficult to invoke explanations based on school differences in implementing the Comer program. This is because the implementation index is as highly correlated with the student social and academic climate constructs as they are with each other. There is little discriminant validity; the implementation index is just another school climate indicator. On reflection, this finding is not very strange, given that Comer's explication of implementation emphasizes how teams operate and relate to a school's mission and how planning procedures and norms of mutual respect operate.

Still, the close correspondence between climate and implementation helps explain why the implementation index also fails to make the Comer effects in Figure 4 go away when it is introduced into the analysis. The present sample contains Chicago schools with better and worse climates, and both students and staff agree on which is which. It also contains schools that vary in how Comer-like staff see them to be. But within limits imposed by studying only 19 schools with many similar characteristics from within a single city, it does not seem that the schools with better-measured program implementation and better or faster changing climates are necessarily the same schools where student achievement and acting out are changing the most. So analyses of individual schools cannot sustain the conclusion suggested by comparing the average Comer and average control

school—that school climate variables mediate between implementing the Comer program and student changes in achievement and acting out.

DISCUSSION

This study of Comer's School Development program took place in inner-city Chicago schools that are almost exclusively minority and attended by African-American children from very poor homes. It also took place at a time when the city schools had begun a highly publicized school reform initiative that was in place for the whole study period. This reform may explain why staff in both program and control schools saw steady improvement in many different aspects of school climate and it implies that, to be effective by the criteria we set, Comer schools had to do better than an escalator that was already moving upwards. These realities of place, population and context frame all the results we discuss below.

Being in an officially designated program school had only a modest impact on staff perceptions of the quality of program implementation and no measured impact on staff perceptions of the school's social climate. However, by the last two study years it did lead both teachers and students to believe that program schools had initiated more pedagogically relevant activities. And as early as the program's first year, students in Comer schools saw their social climate as better. By the end of the study period, students had improved in reading and math by about three points, self-reported acting out had decreased by just under half of a behavior, anger was reported as being less frequent, and beliefs about what constitutes misbehavior were more mainstream.

It is noteworthy that the program and control schools started at about the same position for the non-achievement outcomes, suggesting that the results may well be causal. However, the program schools started out behind the controls in academic achievement and then caught up with them, thus creating differences in slopes rather than 8th grade means. Moreover, the achievement results obtained with the inferentially more powerful longitudinal

design were not replicated with the weaker cross-sectional design. So, making causal inferences about the difference in achievement gains is especially dependent on the design used and the adequacy of the statistical control for selection.

Nonetheless, the data suggest that the Comer Program caused positive changes in two socially important sets of outcomes--standardized test scores, on the one hand, and beliefs, feelings and behaviors relevant to disruptive and illegal behavior, on the other. It is rare for a single study to show changes in each of these two domains that are each so central to current discussions of the need for school reform, particularly in the impacted inner-city areas examined here. If such effects were to be found at other Comer sites, they would have important implications for educational policy.

However, they were not found in the only other experimental study of Comer conducted to date—in Prince George’s County. But in that site middle schoolers were the target of the intervention and not younger students; the students were suburban rather than urban; program exposure was mostly for two years rather than up to six; the facilitators were teachers or counselors rather than social workers; and the program variant did not have the emphasis on academic achievement that came to be important in Chicago (Cook et al, 2000). We cannot be sure which of these many site differences, alone or in combination, account for the different student results by site. But all the presumptions are that the Chicago study is superior. It involves longer exposure to the treatment, a treatment variant that respects academic goals for schools as well as social ones, and the population is the one for which Comer created his program and about which national concern is greatest.

As promising as these initial Chicago results for achievement and acting out seem to be, there is no evidence that the program also improved either mental health or positive social behaviors. Nor is there evidence that achievement and acting out changed in the same Comer schools or even in the same students, though the small number of schools makes these last findings quite uncertain. A model of independent program effects seems more appropriate in which changes in academic performance and social behavior are not causally

linked. And finally, it is not yet possible to adduce evidence about whether the student effects persist as students leave their Comer elementary schools and enter high schools that vary considerably in quality. So, we need to replicate the current findings in other school districts where there is a reasonable presumption of quality program implementation and long exposure periods, and we also need to assess the longer-term Chicago program effects on both high school performance and involvement with the justice system.

Let us turn now to explaining the demonstrated Chicago effects. One possible explanation invokes the quality with which the Comer program was implemented. However, the implementation index failed to make the program effects go away when it was introduced into models of program effects. So, the extent to which a school's ethos is Comer-like is not strongly related to the outcome changes caused by the Comer manipulation. Such a conclusion depends on the validity of the implementation index. The one we used meets all the usual psychometric criteria with respect to reliability and face validity. It was created after writing a paper on the program together with its designer (Anson et al, 1991). It was checked out with both Yale and Prince George's County Comer officials before being used. It is now being used in slightly modified form by Abt Associates in their Detroit evaluation of Comer and in a somewhat more revised form by researchers at the School Development Program headquarters. And each year the quantitative and qualitative researchers independently rank-ordered Chicago Comer schools in terms of implementation quality, achieving results that were always almost identical. The index could be improved, particularly if it placed emphasis on how well the facilitator played his or her role and if it were more behavioral in orientation, specifying concrete actions that had taken place rather than staff judgments and beliefs.

The implementation index turned out to be highly correlated with staff and student measures of the school social and academic climate. So, it is not surprising that adding the climate measures to the causal models also failed to reduce the Comer effects by much. It does not seem, therefore, that either implementation of climate factors (as we have measured

them) mediate between a school entering the Comer program and students subsequently improving in achievement and acting out. However, the failure to discover a mediating role for school climate when individual schools were the unit of analysis was at variance with the results when the average Comer and the average control school were compared to each other. Then, climate did change after treatment assignment, social climate changes preceded academic ones, and it was only when both these changes had occurred that achievement and acting out results could be clearly seen. It is not at all clear why the predicted sequence of program effects was not obtained with one unit of analysis (individual schools) but was with a different unit (averages of treatment schools). The difference may have to do, though, with the fact we only had 19 schools—a better number for assessing program main effects than what are essentially interactions between Comer status and standing on the climate mediating variables.

Another explanation other than implementation and climate as measured involves the length of program exposure. Most of the effects on students are not clear until the last two program years, these being the sixth year for the original pilot schools, the fifth for the original Phase I schools and the fourth for the Phase II schools. Two other factors suggest the explanatory importance of program exposure. Figures 4 and 6 show that results from the longitudinal design are slightly larger and considerably clearer than results from the cross-sectional design. This could be because, by the later assessment period, longitudinal students had been in their schools for about a year longer than the average student in the cross-sectional design. The second type of evidence comes from comparing the Chicago and Prince George's County evaluations of Comer's program. In Maryland, students attended the local middle schools for two years and showed no program effects, the very finding that would have resulted if the Chicago study been restricted to two years of program exposure. (But since there are many other site differences, we cannot be sure that variation in the length of program exposure is alone responsible for site differences in student outcomes. But it might be).

A simple exposure explanation makes most sense when an effective program is implemented that is stable in content. But that was definitely not the case in Chicago. During the study period, the national Comer program was experimenting with new ways to become more instructionally relevant; and Youth Guidance sought to reformulate certain aspects of program design to reflect its own experiences in community organizing, delivering mental health services and relationship-building in Chicago neighborhoods. Thus, the Comer schools probably benefited from more sophistication and enthusiasm about changing the social life in schools than it had received anywhere else before. But even this emphasis waxed and waned. When pointed pedagogic priorities eventually emerged from the school district, Youth Guidance began to work on improving instruction and to reduce the saliency of some of its earlier concerns, particularly around parent involvement and helping principals manage a decentralized school. Of course, a priority shift like this is not problematic for length of exposure explanations if the shift pattern over time mirrors what an effective program theory specifies. Comer recommends that socially disorganized and demoralized schools should first improve their social climate, then improve the academic focus, and then children's behavior and performance will "somehow" improve. As Figures 2 through 6 illustrate, this temporal sequence of changes actually occurred.

Although the sequence Comer specifies seems plausible, it could not be tested sensitively with the data on hand. First, there is the problem that analyses at the school level fail to corroborate the same sequential model. Second, we do not know what would have happened had Youth Guidance begun the program with the pedagogical priorities it ended up with. Would it then have achieved the program effects we demonstrated in later years? Achieving them would imply that changing the school governance structure and social climate are not as necessary for improving academic climate and student outcomes as Comer maintains. But after observing in some of these schools, we find it difficult to believe that any Comer implementation would have been successful had it started with instructional rather than social priorities. These schools were initially characterized by mistrust,

factionalism, resignation and counterproductive staff beliefs about how family life in desperately poor neighborhoods impacts on school life (Bryk et al, 1998). There was also considerable pessimism about the levels of support to be expected from the district central office, from parents, and even from principals and vice-principals who were not all as pedagogically engaged or knowledgeable as they might have been (Payne, 1998).

But working to improve a school's social climate may not by itself be sufficient to bring about important student changes. After four years of program development in Prince George's County the local priority was still on using the Comer program to improve social relationships and disseminate knowledge of child development. Improving instructional practice was not a major program goal, and no student changes occurred in either social or academic outcomes. What happens during program exposure is surely more relevant than the length of exposure per se. We suspect that the best schools are those that begin with activities to improve both a school's social climate and its academic focus. Indeed, the two are not very distinct in either the staff or student view. Initial emphasis exclusively on the social and psychological may sometimes impede the transition to working on pedagogic changes later. Needed is an organization that acknowledges the theoretical and operational interdependence of both factors.

A key to possible explanation has to lie in what Youth Guidance actually did to make a difference. But before describing this we must note that the Chicago Comer effort did not exist in a vacuum and that some background factors are crucial as part of a complete explanatory account of the demonstrated program effects. First, the Comer program was implemented in a Chicago context where reform was the norm and, as our implementation and climate data indicated, many schools were already improving. The fact that these schools included the control schools created a higher hurdle for the Comer schools to jump over if they were to show effects. But it probably also helped the Comer schools in several different ways. Staff did not have to be persuaded of the political necessity to change, of the consequences of not changing and of the importance of working collaboratively in order to

effect change. In addition, the central district office left no ambiguity about the direction of desired change, coming to emphasize academic achievement in all its later explicit pronouncements and in its new policies about school probation and ending student social promotions. Other Comer sites do not enjoy a local context where it is impossible to maintain the status quo and where the goal for change is so clear, even if some disagree with the goal given their distaste for standardized achievement scores.

Second, the institution responsible for the Chicago program, Youth Guidance, has a unique organizational culture that few other agencies implementing the Comer program can replicate, mostly because they are offices within a school district. Youth Guidance has been actively working in inner-city, minority communities for decades. About half of its personnel are African-American, and all are social workers trained and eager to work with the population under study. The organization truly respects parents and understands their difficult lives, as was illustrated in the first year by the overwhelming parent response to Youth Guidance's outreach into the public housing units where nearly all the pilot children lived. The organization also had the interpersonal skills required for supporting and guiding principals without undermining their authority or taking sides in school battles. Also noteworthy was the ability of facilitators to persevere, given they had principals who differed in personal commitment to the program, in their confidence as principals, in their social skills, in their inclination to democracy, in their energy levels and in their knowledge of instructional practice. Then, there was the organization's consistent refusal to blame others when things went wrong. Instead, there was a steady hand on the tiller in an organization where few senior staff turned over and considerable political skill was shown in dealing with schools, district personnel, funders and even researchers. The organization also frankly acknowledged its inability to take the lead on instructional matters, while simultaneously seeking to learn as much as possible as quickly as possible about how to improve classroom teaching. Many other organizations seeking to implement the School

Development Program will not have the political and social skills of Youth Guidance; nor will they enjoy its level of support in local communities.

Some specific activities that Youth Guidance undertook did not turn out to be very helpful in the long term, as the organization itself learned. In the pilot schools' first year, Youth Guidance instituted a community outreach campaign and canvassed many homes in the public housing units serving the schools, showing an unheard-of level of concern for parents and their relationships to schools. Principals and teachers were amazed at the number of parents who turned out for the first social events and by the level of their general concern about school quality. But with the resources available, such intensive community organization could not be maintained either in the pilot schools or when larger numbers of schools later entered the study. At best, modest numbers of parents in the Comer schools remained involved in school governance, in parent groups and in general school support. Although their continued participation represents a long-term parent effect of unknown magnitude, it seems fair to claim that parent involvement was eventually defined more modestly and became increasingly less important to Youth Guidance.

Some of the positive things Youth Guidance did to bring about effects are the product of its background in social work and the emphasis it places on ways to promote better relationships between staff, to deal with students who are acting out and to support principals in institution-building. As late as the last study year Youth Guidance was still introducing some social programs into schools, including one on peer mediation that was aimed at decreasing anger and improving interpersonal behavior, including violence and other forms of acting out. Much earlier, most of the Comer schools had set up Discovery Rooms where disruptive students could be sent and professional help made available to them. In addition, the Youth Guidance facilitators were widely known and often available in each school, deliberately modeling what it takes to be sensitive in interpersonal relationships yet focused on practical problem solving. Indeed, they experienced great stress in mediating between the various parties within a school, for while climates were improving they were still

far from perfect when this study ended. Some important role seems to have been played by the social work background of Youth Guidance, by the example facilitators set, by the Discovery Room and by the peer mediation program. However, it is not clear whether they functioned as unique causes or as contributory causes whose influence was conditioned by the general improvements in school climate.

We have no comparable details on the control schools, but it is unlikely they would have had social workers playing the hands-on roles Youth Guidance facilitators played. Traditional school social workers do not implement Comer-like programs! It is also unlikely that the control schools would have had specific analogs to a Discovery Room or to the peer mediation implemented in Comer schools. Nor would the control schools have had in-service training about interpersonal relationships like Youth Guidance occasionally provided to the Comer schools. Youth Guidance never gave up trying to maintain a positive school social climate and, even when the pressure was on them to demonstrate achievement gains, they were still introducing some new programs to enhance the social climate and improve students' social behavior.

As achievement gains became more politically salient, Youth Guidance needed to respond. It did so, no longer being required to enroll new schools. It began by hiring its own director of instruction and by developing links with a faculty member in a local university who had developed an instructional package. However, neither worked out. The director of instruction was hired away after a year; and schools could provide very little staff development time for learning the instructional package. Teacher volunteering on weekends was mostly required. Anyway, the package developers were spread very thinly across the city schools with which they had a contract, and Youth Guidance made the decision to adopt the package without consulting all the relevant personnel in all the Comer schools. But Youth Guidance learned from these mistakes, being a self-reflective organization that held weekly staff meetings to which outside experts were sometimes invited, including members of the Yale Comer staff.

When Youth Guidance's instructional efforts became more focussed, they were still not as organized as the efforts directed at improving school social climate, in part because the focus on staff relationships, governance and difficult students could not be abandoned. Eventually, though, a set of pedagogic services evolved that were responsive to district and principal needs. Salient among these were: Facilitating the transfer of some poor teachers, coordinating meetings between various Comer schools in order that teachers could learn from each other, increasing the amount of within-school dialogue about instruction, trying to get principals to observe in classrooms and provide feedback, linking some less effective classroom teachers to master teachers, and providing schools with information about instructional packages in particular curriculum areas. Youth Guidance never provided a specific instructional package and, although the Yale authorities have flirted with some program developers in the past, they are also not yet ready to prescribe a specific model. Indeed, it is in the spirit of the program not to be too prescriptive, instead suggesting a range of vetted packages for schools to consider, or coaching staff about the criteria for judging among packages, or helping staff get feedback about any package they are considering. However, given the late start with curriculum matters in Chicago and the process orientation that characterizes the program at large, it is understandable that the Chicago "instructional" components are more difficult to pin down than the social components.

The Comer program is not very prescriptive about school goals and strategies. So long as certain structures and traditions of decision-making and collaboration are in place, schools can do pretty much what they want in terms of goals and strategies for reaching them. So, Youth Guidance facilitators were needed to set up the teams and to educate staff to the traditions of decision making and collaboration, based partly on Comer's guidelines, partly on their own training as social workers, and partly on their reflections on experience in the schools. This is why they suggested the use, not only of long-established parts of Comer's program like Discovery Rooms, but also the use of peer mediation that is not usually a part of the program. They particularly felt called upon to be assertive with

principals, coaching them about their social relationships with staff and parents and also about their relationships with the central office and local school councils and about school plans that might be evolved. Youth Guidance central office staff also felt the need to canvass the country to find out about effective and practical curricula and instructional techniques that might meet the needs of local schools. Youth Guidance staff were rarely advocates; but they actively provided options to schools that could not readily generate their own. And they kept persisting until decisions that had to be made were in fact made.

This raises the question: How much was this a Comer effect and how much a Youth Guidance effect? Throughout the implementation period, Youth Guidance officials were in close contact with the School Development Program headquarters in Yale, learning from their colleagues there but also teaching them. On very many occasions Comer's theoretical work clearly provided guidelines for practical actions, serving both as a blueprint and a general orienting device. So, Comer is intrinsic to what happened. Even so, this study is more an evaluation of Youth Guidance's variant of the Comer program than it is an evaluation of the program per se, given the way some implementation specifics are supposed to vary by school district. This may be one reason why the Chicago results were so different from those achieved by the local program variant in Prince George's County. Comer's is no cookie cutter whole school program, as its emphasis on process guarantees. So, the bottom line on Comer's program has to depend on synthesizing results across several school districts that implement the program with different priorities and practices.

REFERENCES

- Angrist, J.D., Imbens, G.W. & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444-462.
- Anson, A., Cook, T.D., Habib, F., Grady, M.K., Haynes, N. & Comer, J.P. (1991). The Comer School Development Program: A Theoretical Analysis. *Journal of Urban Education*, 26, 56-82.
- Becker, B.J. & Hedges, L.V. (1992). *A Review of the literature on the effectiveness of Comer's School Development Program*. New York: The Rockefeller Foundation.
- Bryk, A.S., Lee, V.E. & Holland, P.B. (1993). *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.
- Bryk, A.S., Raudenbush, S.W. & Congdon, R.T. (1996). *Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. Chicago: Scientific Software International
- Bryk, A.S., Sebring, P.B., Kerbow, D., Rollow, S. & Easton, J.Q. (1998). *Charting Chicago School Reform: Democratic Localism as a Lever for Change*. Boulder, Colorado: Westview Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Social Sciences (2nd Ed.)* Hillsdale, N.J.: Lawrence Erlbaum.
- Comer, J.P. (1980). *School Power*. New York: The Free Press
- Comer, J.P. (1984). Home-school relationships as they affect the academic success of children. *Education and Urban Society*, 16 (3). 323-337.
- Comer, J.P. (1988). Educating poor minority children. *Scientific American*, 259 (5). 42-48.
- Comer, J.P., Haynes, N.M., Joyner, E.T. & Ben Avie, M. (1996). *Rallying the Whole Village: The Comer Process for Reforming Education*. New York: Teachers College Press.
- Cook, T.D., Habib, F.N., Phillips, M., Settersten, R.A., Shagle, S.C. & Degirmencioglu, S.M. (2000). Comer's School Development Program in Prince George's County, Maryland: A Theory-Based Evaluation. *American Educational Research Journal*.
- Ferguson, R.F. (1998). Can Schools Narrow the Black-White Test Score Gap? In C. Jencks & M. Phillips. *The Black-White Test Score Gap*. Washington, DC: Brookings. 318-374.
- Finn, J.D. & Achilles, C.M. (1990). Answers and questions about class size. A statewide experiment. *American Educational Research Journal*, 27(3). 557-577.

- Haynes, N.M., Emmons, C.L., Gebreyesus, S. & Ben-Avie, M. (1996). The School Development Program evaluation process. In J.P. Comer, N.M. Haynes, E.T. Joyner & M. Ben-Avie (Eds.) *Rallying the Whole Village: The Comer Process for Reforming Education*. New York: Teachers College Press.
- Maclver, D.J., Reuman, D.A. & Main, S.R. (1995). Social structuring of the school. In M.R. Rosenzweig and L.W. Porter (Eds.) *Annual Review of Psychology*, 46, 375-440. Palo Alto, CA: Annual Reviews, Inc.
- McLaughlin, M.W. & Oberman, I. (Eds.) (1996). *Teacher learning: New policies, new practices*. New York: Teachers College Press.
- Millsap, M., Chase A., Cook, T.D. & Hunt, H. D. (In Preparation). *School Development Program in Detroit*.
- Mosteller, F., Light, R.J. & Sachs, J.A. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Educational Review*, 66, 797-842.
- Payne, C. (1998). *Overview of the Chicago Comer Ethnography*. In Preparation.
- Phillips, M. (1997). What makes schools effective: A comparison of the relationships of communitarian climate and academic climate mathematics achievement and attendance during middle school. *American Educational Research Journal*, 34(4), 633-662.
- Rosenthal, R., & Rubin, D.B. (1994). The counternull value of an effect size: A new statistic. *Psychological Science*, 5, 329-334.
- Rutter, M., Maugham, B., Mortimore, P. & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.
- Squires, D.A. & Joyner, E.T. (1996). Time and alignment: Potent tools for improving achievement. In Comer, et al. (Eds.) *Rallying the Whole Village: The Comer Process for Reforming Education*. New York: Teachers College Press.
- Wilson, W.J. (1987). *The truly disadvantaged: The inner city, the underclass and public policy*. Chicago: University of Chicago Press.
- Wilson, W.J. (1996). *When work disappears: The world of the new urban poor*. New York: Alfred A. Knopf.

Table 1: Student Climate and Outcome Measures

| Measure | Number of Items | Average Reliability | Mean | Std. Dev. | Sample Item and Question Stem Where Appropriate |
|---|------------------------|----------------------------|-------------|------------------|---|
| Social Climate | | | | | |
| Social Relationships with Adults in School | | | | | |
| Perceived School Staff Respect for Students | 6 | .75 | 3.56 | 0.18 | How many of your teachers really listen to what you have to say? |
| Perceived Caring of School Staff | 3 | .87 | 3.65 | 0.19 | When you have a problem, how often can you depend on...your teachers to help you out? |
| Social Climate Among Students | | | | | |
| Social Skills of Students | 9 | .87 | 3.23 | 0.19 | How many students in this school...accept criticism without getting angry? |
| Use of Positive Problem Solving Strategies | 3 | .75 | 2.83 | 0.26 | When problems occur among students in this school, how often do students...work out the problems amongst themselves by talking it over? |
| Use of Negative Problem Solving Strategies (R) | 4 | .73 | 2.60 | 0.20 | When problems occur among students in this school, how often do students work out the problems by fighting, hitting, either at or away from school? (R) |
| Perception of Physical Safety | 4 | .77 | 3.63 | 0.27 | How often do you worry about...getting beaten up at school? (R) |
| Pride in the School | 6 | .85 | 3.23 | 0.27 | How proud do you feel of this school? |
| Attachment to the School | 6 | .87 | 3.26 | 0.30 | If you had to stop going to this school, how much would you miss...your fellow students? |
| Comfort in Interpersonal Contexts | 8 | .85 | 3.51 | 0.18 | How comfortable do you feel...talking to your teacher in class? |
| Personal Feelings of Not Belonging in School (R) | 5 | .80 | 3.61 | 0.18 | I feel like no one knows who I am in this school. |
| Academic Climate | | | | | |
| Academic Relationships with Adults in School | | | | | |
| Teachers' Encouragement of Academic Success | 5 | .79 | 3.99 | 0.17 | Of the teachers you know in this school, how many...tell students that doing well in school now is important to their future? |
| Teachers' Concern with Childrens' Learning Behaviors to Motivate Students | 6 | .71 | 3.99 | 0.17 | How many of your teachers...don't care if you get bad grades? (R) |
| | 10 | .80 | 3.70 | 0.17 | How many of the teachers in this school are willing to go out of their way and spend extra time to help students do their best? |
| Recognition for Achievement | 4 | .85 | 3.69 | 0.29 | How many times are students recognized or given awards at school assemblies or other special gatherings for... making good grades? |
| Academic Values Among Students | | | | | |
| Student Devaluing of Academic Success (R) | 4 | .64 | 3.42 | 0.16 | In this school, kids who get good grades are called names. (R) |
| Student Acceptance of School Values | 7 | .79 | 3.46 | 0.16 | How many students in your school think doing homework is important? |
| Personal Valuing of Education | 3 | .69 | 3.19 | 0.17 | School is boring. (R) |

Table 1: Student Climate and Outcome Measures

Student Outcomes

Mental Health Attributes

| | | | | | |
|---|---|-----|------|------|--|
| Self-Efficacy in School | 9 | .78 | 3.78 | 0.15 | How well can you remember things taught in class and school books? |
| Racial Pride | 5 | .81 | 3.97 | 0.16 | How proud are you of what that group has done? |
| Positive Expectations for Future Life | 6 | .77 | 4.20 | 0.16 | What are the chances that someday you yourself will...stay free of drugs? |
| Valuing of Positive Outcomes in Future Life | 6 | .74 | 3.92 | 0.18 | How important is it to you that you...will go to church? |
| Satisfaction with Life and Relationships | 9 | .80 | 4.01 | 0.15 | How happy are you with...the kind of person you are? |
| Use of Constructive Coping Strategies | 3 | .68 | 4.15 | 0.15 | When you don't do well on schoolwork, how often do you...tell yourself you will do better next time? |
| Lack of Depression | 5 | .79 | 3.85 | 0.23 | In the last year, how often have you felt...hopeless about the future? |
| Lack of Anger | 4 | .81 | 3.18 | 0.24 | In the last year, how often have you felt...really mad at other people? |
| Use of Passive Coping Strategies (R) | 5 | .72 | 4.15 | 0.15 | When you don't do well on schoolwork, how often do you...say it was the teacher's fault? |

Negative Social Behaviors

| | | | | | |
|----------------------------|----|-----|------|------|--|
| Acting Out | 11 | .87 | 3.31 | 2.94 | In the past year, how often have you...done some pretty dangerous things just for fun? |
| Disapproval of Misbehavior | 8 | .88 | 4.01 | 0.22 | How wrong is it...to hit someone if they mess with someone in your family? |
| Substance Abuse | 4 | .67 | 0.53 | 0.25 | In the past month or 30 days, how many cigarettes did you smoke? |

Positive Social Behaviors

| | | | | | |
|---|----|-----|------|------|---|
| Time Spent in Clubs or Lessons | 6 | .87 | 0.16 | 0.04 | On a weekend, how many hours do you spend taking part in school clubs or groups? |
| Time Spent Playing Sports | 4 | .79 | 0.22 | 0.03 | On a school day, how many hours do you spend playing pickup games outside of school like basketball, softball, etc? |
| Time Spent Doing Homework or Free Reading | 4 | .76 | 0.25 | 0.04 | On a weekend, how many hours do you spend reading for fun? |
| Time Spent in Unstructured Activities | 10 | .87 | 0.38 | 0.05 | On a weekend, how many hours do you spend talking on the telephone? |

Academic Achievement

| | | | | | |
|------------------------|-----|-----|-------|------|---------------------------|
| Math NCE Test Score | N/A | N/A | 32.00 | 7.09 | ITBS math score -- NCE |
| Reading NCE Test Score | N/A | N/A | 33.19 | 4.57 | ITBS reading score -- NCE |

Family Process

| | | | | | |
|---------------------------------|---|-----|------|------|--|
| Home Academic Support | 5 | .72 | 3.00 | 0.17 | How often does someone at home...look to make sure you've done your homework? |
| Parenting Style: Reasoning | 3 | .75 | 4.10 | 0.14 | When you break one of your parents more important rules, how often do they...ask you to explain? |
| Parental Rule Setting | 6 | .66 | 3.81 | 0.14 | In your home, how often do you have to follow rules about...what time you go to bed at night? |
| Positive Parent-Child Relations | 6 | .81 | 3.97 | 0.14 | My parents...respect my ability to make decisions. |

(R) = Reverse Coded

Table 2: Comparability of Comer and Control Schools

| School Characteristics: Mean of Four Years (Source: District Archives) | | | | | | | | | |
|--|------------|-------------------|--------------|--------------|--------------------|---------|---------|---------|------------|
| | Enrollment | Limited English % | Low Income % | Attendance % | Student Mobility % | White % | Black % | Asian % | Hispanic % |
| Control | 676 | 10 | 94 | 92 | 31 | 3 | 74 | 5 | 18 |
| Comer | 648 | 11 | 91 | 92 | 30 | 1 | 81 | 4 | 13 |

| Student and Family Demographics: Mode for Four Years (Source: Student Self-Report) | | | | | | | | | |
|--|--------|--------------------------|----------------|---------------------------|-------------|-------------|------------------------|-----------------------|-------------------------|
| | Sex | Highest Parent Education | | Highest Parent Employment | | | Family Structure | | |
| | Male % | HS or Less % | Some College % | Not Working % | Part Time % | Full Time % | 2 Biological Parents % | 1 Biological Parent % | No Biological Parents % |
| Control | 49 | 53 | 47 | 22 | 13 | 65 | 35 | 46 | 19 |
| Comer | 45 | 54 | 46 | 20 | 19 | 62 | 29 | 45 | 27 |

| Selected Outcome Variables: First Study Year (Source: District Archives and Student Self-Report) | | | | | | | | | | |
|--|-----------------------|----------------------------|-----------------------|---------------------------------|-------------|---------|------------------|----------------------------|-------------------------|--------------------|
| | Home Academic Support | Family Process | | | Achievement | | Student Outcomes | | | |
| | | Parenting Style: Reasoning | Parental Rule Setting | Positive Parent-Child Relations | Math | Reading | Lack of Anger | Disapproval of Misbehavior | Self-Efficacy in School | Lack of Depression |
| Control | | | | | | | | | | |
| Mean | 3.13 | 4.17 | 3.95 | 3.97 | 34.76 | 36.30 | 3.05 | 4.09 | 3.81 | 3.63 |
| Std. Dev. | 0.16 | 0.15 | 0.09 | 0.12 | 7.68 | 3.86 | 0.15 | 0.13 | 0.18 | 0.12 |
| Comer | | | | | | | | | | |
| Mean | 3.10 | 4.05 | 3.93 | 3.93 | 29.07 | 33.08 | 3.05 | 4.03 | 3.77 | 3.58 |
| Std. Dev. | 0.11 | 0.11 | 0.09 | 0.08 | 6.24 | 4.23 | 0.14 | 0.18 | 0.15 | 0.10 |

Table 3: Staff Implementation and Climate Measures^a

| Measure | Number of Items | Average Reliability | Mean | Std. Dev. | Items and Sample Questions |
|--|-----------------|---------------------|------|-----------|--|
| Decentralized School Management | | | | | |
| Efforts to Implement Decentralized Mgmt. | 7 | .89 | 3.60 | 0.22 | There is an active effort in my school to...involve more groups by consensus (agreement) rather than by vote. |
| Success in Implementing Decentralized Mgmt. | 8 | .90 | 3.35 | 0.26 | To what extent do all members of the school community (teachers, parents, non-teaching staff, and administrators) have a say in the decision-making bodies of your school? |
| Voluntary Involvement in School Governance | 4 | .83 | 2.62 | 0.22 | Since the beginning of the school year, how often have you had input into...committees that plan or influence school policy or programs? |
| Quality of Structures for Decentralized Mgmt. | 5 | .89 | 2.75 | 0.32 | How effective is your school's team that deals with student behavioral problems and child development issues? |
| Open Communication Channels | 3 | .81 | 3.50 | 0.30 | How free do teachers and other staff members feel to communicate with the principal? |
| Staff Open to Innovative Ideas | 3 | .90 | 3.38 | 0.25 | To what extent does the staff in your school try to...identify new ways of doing things? |
| Staff Committed to Change | 8 | .74 | 3.64 | 0.10 | I am willing to put in extra effort to help the current school change process be successful. |
| School Social Climate | | | | | |
| Teacher Responsiveness to Students ^c | 7 | .91 | 3.59 | 0.23 | How many instructional staff in your school...make special efforts to get to know students? |
| Quality of Relations Within School | 6 | .86 | 3.58 | 0.23 | How would you characterize the quality of...teacher/teacher relationships? |
| Prevalence of a No-Fault Orientation | 6 | .85 | 2.89 | 0.25 | When a problem arises in your school...adults usually blame each other rather |
| Teachers' Pride in the School ^b | 3 | .84 | 3.61 | 0.35 | How many of the instructional staff in your school...feel pride in the school and in its students? |
| Non-Teaching Staffs' Pride in the School | 3 | .85 | 3.34 | 0.32 | How many adults who work in your school who are NOT teachers...feel pride in this school and in its students? |
| General Attachment to the School | 4 | .90 | 2.90 | 0.44 | If you had to stop working at this school, how much would you miss...the approach to education? |
| Emphasis on Positive Multiculturalism | 5 | .71 | 3.72 | 0.26 | To what extent...is there tension between school staff members of different racial or ethnic backgrounds? |
| Quality of School Rules | 3 | .83 | 3.25 | 0.36 | The rules in my school are...clearly stated. |
| Interpersonal Social Skills of Students | 8 | .93 | 3.05 | 0.31 | In your school, how many students...cooperate with other students? |
| School-Related Behavior of Students | 3 | .81 | 3.72 | 0.24 | How many students in your school have their progress hurt because of their...acting out in school? |
| School Academic Climate | | | | | |
| Teachers' Extra Effort on Behalf of Students ^c | 4 | .81 | 3.39 | 0.24 | To motivate students, how often do you...provide opportunities for students to select their own projects? |
| Response to School Problems ^b | 4 | .64 | 2.98 | 0.10 | When students skip class how often do you...talk to the student about why they were absent? |
| Teachers Feel Efficacious ^b | 7 | .93 | 3.72 | 0.21 | How much can you do to...assist parents in helping their children do well in school? |
| Academic Motivation Attributed to Students | 5 | .87 | 2.88 | 0.27 | How many students in your school...are willing and eager to learn? |
| Future Academic Expectations for Students | 3 | .74 | 3.29 | 0.25 | How many students in your school...do you expect to complete high school? |
| Parental Involvement | | | | | |
| Parent Involvement in General School Activities ^b | 5 | .81 | 2.11 | 0.14 | How many of the students you teach have parents who...help out in your classroom with academic work? |
| Parent Valuing of Education | 5 | .89 | 2.85 | 0.28 | In your judgment, how many parents or guardians of students in your school...expect their children to complete a four-year college degree? |
| Interaction with Parents about their Children ^c | 5 | .83 | 3.12 | 0.20 | How many of your students' parents have you...met with at regularly scheduled conferences intended for all students? |

^a All items used a scale of 1 ("not at all/none/nothing/strongly disagree") to 5 ("a great deal/very effective/all/strongly agree").

^b Only Teachers Responded to the Items in these Measures

Table 4: Comer Effects on Implementation Means and Slopes - School and Individual Level Analyses

| Measures | Mean | Std. Dev. | School Level Analysis | | | | Individual Level Analysis | | | |
|---|------|-----------|--|------|----------------------|------|--|------|----------------------|------|
| | | | Differences in Mean Level ^a | | Differences in Slope | | Differences in Mean Level ^a | | Differences in Slope | |
| | | | b | se | b | se | b | se | b | se |
| Overall Implementation | 3.06 | 0.31 | .06 | .108 | .01 | .038 | .01 | .012 | .00 | .012 |
| Every School has an LSC. Most also have one other team responsible for planning and management of school-wide activities. How effective is this team in your school, however it is named? (SPMT question) | 2.90 | 0.39 | .15 | .133 | .05 | .042 | .01 | .019 | .02 | .019 |
| How effective is your school's team that deals with student behavioral problems and child development issues? (SST question) | 2.68 | 0.48 | .06 | .238 | .09 | .065 | -.06** | .019 | .04* | .019 |
| How effective is your parent group? (PTA question) | 2.57 | 0.41 | .02 | .133 | -.01 | .055 | .02 | .018 | -.00 | .018 |
| How effective is the School Improvement Plan? | 3.05 | 0.43 | .07 | .140 | .06 | .056 | -.02 | .018 | .03 | .018 |
| How effective is communication between teams in your school so that each team knows what the others are doing? | 2.58 | 0.34 | .07 | .131 | -.01 | .061 | .04* | .018 | -.01 | .018 |
| To what extent is knowledge of child or adolescent development used in developing programs and problem-solving strategies in your school? | 3.24 | 0.33 | .04 | .114 | .02 | .038 | .01 | .017 | .00 | .017 |
| To what extent are decisions about school policy and goals made by consensus (group agreement) rather than by vote? | 3.26 | 0.36 | .25* | .104 | .00 | .034 | .11*** | .017 | .01 | .017 |
| To what extent are team members really committed to the school programs aimed at improving your school? | 3.56 | 0.31 | -.02 | .117 | -.01 | .042 | .01 | .015 | -.01 | .015 |
| To what extent do all members of the school community (teachers, parents, non-teaching staff and administrators) have a say in the decision-making bodies of your school? | 3.26 | 0.32 | .14 | .109 | -.02 | .039 | .08*** | .017 | -.02 | .017 |
| To what extent are the special concerns of the various cultural and racial groups given attention in your school? | 3.42 | 0.33 | -.13 | .123 | -.05 | .038 | -.03* | .017 | -.04* | .017 |
| Decentralized School Management | | | | | | | | | | |
| Efforts to Implement Decentralized Mgmt. | 3.60 | 0.22 | .09 | .069 | -.03 | .025 | .07*** | .012 | -.02 | .012 |
| Success in Implementing Decentralized Mgmt. | 3.35 | 0.26 | .03 | .087 | -.02 | .030 | .03* | .011 | -.02 | .012 |
| Voluntary Involvement in School Governance | 2.62 | 0.22 | .15+ | .077 | -.03 | .034 | .09*** | .014 | -.02 | .014 |
| Quality of Structures for Decentralized Mgmt. | 2.75 | 0.32 | .07 | .114 | .04 | .040 | -.00 | .013 | .02 | .014 |
| Openness to Change | | | | | | | | | | |
| Open Communication Channels | 3.50 | 0.30 | -.00 | .133 | -.04 | .036 | .03 | .013 | -.02 | .013 |
| Staff Open to Innovative Ideas | 3.38 | 0.25 | .05 | .079 | .02 | .029 | .01 | .013 | .03 | .013 |
| Staff Committed to Change | 3.64 | 0.10 | .00 | .037 | -.01 | .016 | .01 | .007 | -.01 | .007 |

+ = p<.10 * = p<.05 ** = p<.01 *** = p<.001

^a Estimated at the final time point

^v Estimated at the midpoint of the time period

Table 5: Implementation Means by Year and Comer Status

| Measure | YEAR | CONTROL (N=9) | | | | COMER (N=10) | | | |
|---|-----------|---------------|------|------|------|--------------|------|------|------|
| | | 2 | 3 | 4 | 5 | 2 | 3 | 4 | 5 |
| Overall Implementation | Mean | 2.93 | 2.98 | 3.05 | 3.12 | 2.97 | 3.09 | 3.09 | 3.21 |
| | Std. Dev. | 0.29 | 0.43 | 0.29 | 0.23 | 0.23 | 0.30 | 0.33 | 0.38 |
| Every School has an LSC. Most also have one other team responsible for planning and management of school-wide activities. How effective is this team in your school, however it is named? (SPMT question) | Mean | 2.84 | 2.78 | 2.88 | 2.91 | 2.82 | 2.96 | 2.92 | 3.09 |
| | Std. Dev. | 0.37 | 0.49 | 0.38 | 0.37 | 0.32 | 0.42 | 0.41 | 0.42 |
| How effective is your school's team that deals with student behavioral problems and child development issues? (SST question) | Mean | 2.68 | 2.64 | 2.80 | 2.71 | 2.49 | 2.61 | 2.71 | 2.81 |
| | Std. Dev. | 0.49 | 0.50 | 0.43 | 0.44 | 0.42 | 0.52 | 0.45 | 0.67 |
| How effective is your parent group? (PTA question) | Mean | 2.37 | 2.54 | 2.50 | 2.71 | 2.47 | 2.57 | 2.57 | 2.79 |
| | Std. Dev. | 0.49 | 0.38 | 0.34 | 0.37 | 0.37 | 0.34 | 0.48 | 0.43 |
| How effective is the School Improvement Plan? | Mean | 2.93 | 2.95 | 3.06 | 3.22 | 2.81 | 3.02 | 3.12 | 3.26 |
| | Std. Dev. | 0.40 | 0.60 | 0.33 | 0.34 | 0.41 | 0.42 | 0.42 | 0.42 |
| How effective is communication between teams in your school so that each team knows what the others are doing? | Mean | 2.42 | 2.49 | 2.50 | 2.67 | 2.58 | 2.60 | 2.59 | 2.78 |
| | Std. Dev. | 0.32 | 0.43 | 0.27 | 0.30 | 0.35 | 0.28 | 0.25 | 0.46 |
| To what extent is knowledge of child or adolescent development used in developing programs and problem-solving strategies in your school? | Mean | 3.10 | 3.23 | 3.19 | 3.35 | 3.11 | 3.26 | 3.28 | 3.39 |
| | Std. Dev. | 0.34 | 0.50 | 0.33 | 0.21 | 0.20 | 0.30 | 0.34 | 0.39 |
| To what extent are decisions about school policy and goals made by consensus (group agreement) rather than by vote? | Mean | 3.07 | 3.01 | 3.21 | 3.19 | 3.26 | 3.44 | 3.37 | 3.46 |
| | Std. Dev. | 0.26 | 0.49 | 0.35 | 0.25 | 0.27 | 0.35 | 0.37 | 0.36 |
| To what extent are team members really committed to the school programs aimed at improving your school? | Mean | 3.43 | 3.49 | 3.57 | 3.68 | 3.49 | 3.52 | 3.59 | 3.70 |
| | Std. Dev. | 0.28 | 0.34 | 0.30 | 0.22 | 0.25 | 0.28 | 0.39 | 0.41 |
| To what extent do all members of the school community (teachers, parents, non-teaching staff and administrators) have a say in the decision-making bodies of your school? | Mean | 3.09 | 3.12 | 3.18 | 3.23 | 3.29 | 3.38 | 3.33 | 3.38 |
| | Std. Dev. | 0.35 | 0.44 | 0.23 | 0.26 | 0.26 | 0.25 | 0.34 | 0.33 |
| To what extent are the special concerns of the various cultural and racial groups given attention in your school? | Mean | 3.32 | 3.43 | 3.50 | 3.49 | 3.41 | 3.43 | 3.36 | 3.43 |
| | Std. Dev. | 0.29 | 0.46 | 0.39 | 0.23 | 0.22 | 0.37 | 0.28 | 0.41 |
| Decentralized School Management | | | | | | | | | |
| Efforts to Implement Decentralized Management | Mean | 3.47 | 3.47 | 3.52 | 3.60 | 3.68 | 3.64 | 3.66 | 3.70 |
| | Std. Dev. | 0.24 | 0.28 | 0.18 | 0.19 | 0.17 | 0.19 | 0.22 | 0.24 |
| Success in Implementing Decentralized Management | Mean | 3.20 | 3.26 | 3.34 | 3.42 | 3.31 | 3.39 | 3.39 | 3.47 |
| | Std. Dev. | 0.24 | 0.39 | 0.25 | 0.16 | 0.19 | 0.23 | 0.28 | 0.30 |
| Voluntary Involvement in School Governance | Mean | 2.38 | 2.48 | 2.54 | 2.66 | 2.62 | 2.70 | 2.71 | 2.81 |
| | Std. Dev. | 0.23 | 0.19 | 0.17 | 0.16 | 0.20 | 0.16 | 0.17 | 0.19 |
| Quality of Structures for Decentralized Management | Mean | 2.64 | 2.69 | 2.73 | 2.85 | 2.62 | 2.74 | 2.77 | 2.94 |
| | Std. Dev. | 0.32 | 0.41 | 0.26 | 0.27 | 0.27 | 0.31 | 0.32 | 0.39 |
| Openness to Change | | | | | | | | | |
| Open Communication Channels | Mean | 3.37 | 3.45 | 3.53 | 3.48 | 3.50 | 3.60 | 3.50 | 3.54 |
| | Std. Dev. | 0.24 | 0.45 | 0.28 | 0.24 | 0.27 | 0.30 | 0.29 | 0.32 |
| Staff Open to Innovative Ideas | Mean | 3.25 | 3.33 | 3.39 | 3.45 | 3.25 | 3.39 | 3.44 | 3.52 |
| | Std. Dev. | 0.19 | 0.29 | 0.24 | 0.16 | 0.21 | 0.27 | 0.28 | 0.27 |
| Staff Committed to Change | Mean | 3.60 | 3.58 | 3.67 | 3.66 | 3.63 | 3.66 | 3.67 | 3.66 |
| | Std. Dev. | 0.11 | 0.13 | 0.06 | 0.08 | 0.14 | 0.08 | 0.09 | 0.11 |

Table 6: Comer Effects on Staff Climate Means and Slopes - School and Individual Level Analyses

| Measures | School Level Analysis ^a | | | | Individual Level Analysis | | | |
|--|------------------------------------|------|----------------------|------|---------------------------|------|----------------------|------|
| | Differences in Mean Level | | Differences in Slope | | Differences in Mean Level | | Differences in Slope | |
| | b | se | b | se | b | se | b | se |
| School Social Climate | | | | | | | | |
| Teacher Responsiveness to Students ^b | -.07 | .065 | -.02 | .030 | -.01 | .015 | -.01 | .015 |
| Quality of Relations Within School | -.05 | .080 | .00 | .025 | -.03*** | .009 | -.01 | .009 |
| Prevalence of a No-Fault Orientation | -.01 | .092 | .01 | .031 | -.03* | .012 | .00 | .012 |
| Teachers' Pride in the School ^b | -.08 | .118 | -.01 | .039 | -.02 | .016 | -.01 | .016 |
| Non-Teaching Staffs' Pride in the School | -.13 | .129 | -.03 | .034 | -.04** | .013 | -.03 | .013 |
| General Attachment to the School | -.13 | .156 | -.01 | .043 | -.06*** | .017 | -.01 | .017 |
| Emphasis on Positive Multiculturalism | -.08 | .088 | .02 | .017 | -.06*** | .009 | .01 | .009 |
| Quality of School Rules | -.19 | .142 | -.00 | .059 | -.09*** | .014 | -.01 | .014 |
| Interpersonal Social Skills of Students | -.11 | .136 | .04 | .038 | -.09*** | .012 | .02 | .012 |
| School-Related Behavior of Students | -.11+ | .057 | .01 | .026 | -.06*** | .012 | .01 | .012 |
| School Academic Climate | | | | | | | | |
| Teachers' Extra Effort on Behalf of Students ^b | .08 | .064 | .05 | .030 | .01 | .015 | .01 | .015 |
| Response to School Problems ^b | .04 | .026 | .03 | .019 | -.00 | .012 | .01 | .012 |
| Teachers Feel Efficacious ^b | -.06 | .076 | .04 | .027 | -.04** | .014 | .01 | .014 |
| Academic Motivation Attributed to Students | -.09 | .099 | .03 | .030 | -.06*** | .010 | .01 | .010 |
| Future Academic Expectations for Students | -.05 | .090 | .03 | .028 | -.04*** | .008 | .01 | .008 |
| Parental Involvement | | | | | | | | |
| Parent Involvement in General School Activities ^b | -.07 | .058 | -.01 | .026 | -.02 | .011 | -.01 | .012 |
| Parent Valuing of Education | -.10 | .109 | .01 | .024 | -.05*** | .012 | -.01 | .012 |
| Interaction with Parents about their Children ^b | .07 | .065 | .04+ | .020 | .03 | .016 | .02 | .016 |

+ = p<.10 * = p<.05 ** = p<.01 *** = p<.001

^a (Due to differences in statistical power between the school- and individual-level analyses,

p<.10 is reported only for the school-level)

^b Only Teachers Responded to the Items in These Measures

Table 7: Staff Climate Means by Year and Comer Status - School Level

| Measure | YEAR | CONTROL (N=9) | | | | COMER (N=10) | | | |
|--|-----------|---------------|------|------|------|--------------|------|------|------|
| | | 2 | 3 | 4 | 5 | 2 | 3 | 4 | 5 |
| School Social Climate | | | | | | | | | |
| Teacher Responsiveness to Students ^a | Mean | 3.48 | 3.59 | 3.65 | 3.66 | 3.55 | 3.48 | 3.66 | 3.62 |
| | Std. Dev. | 0.19 | 0.21 | 0.21 | 0.10 | 0.24 | 0.27 | 0.27 | 0.26 |
| Quality of Relations Within School | Mean | 3.49 | 3.58 | 3.67 | 3.67 | 3.48 | 3.53 | 3.63 | 3.64 |
| | Std. Dev. | 0.21 | 0.27 | 0.20 | 0.18 | 0.20 | 0.25 | 0.27 | 0.27 |
| Prevalence of a No-Fault Orientation | Mean | 2.85 | 2.89 | 2.95 | 2.88 | 2.84 | 2.87 | 2.91 | 2.92 |
| | Std. Dev. | 0.27 | 0.23 | 0.17 | 0.17 | 0.20 | 0.29 | 0.29 | 0.36 |
| Teachers' Pride in the School ^a | Mean | 3.50 | 3.63 | 3.63 | 3.72 | 3.51 | 3.57 | 3.68 | 3.65 |
| | Std. Dev. | 0.32 | 0.40 | 0.31 | 0.24 | 0.30 | 0.36 | 0.45 | 0.44 |
| Non-Teaching Staffs' Pride in the School | Mean | 3.21 | 3.33 | 3.43 | 3.47 | 3.24 | 3.30 | 3.33 | 3.40 |
| | Std. Dev. | 0.26 | 0.39 | 0.31 | 0.25 | 0.29 | 0.30 | 0.34 | 0.44 |
| General Attachment to the School | Mean | 2.81 | 2.92 | 2.96 | 3.04 | 2.77 | 2.87 | 2.89 | 2.95 |
| | Std. Dev. | 0.37 | 0.51 | 0.33 | 0.35 | 0.34 | 0.55 | 0.52 | 0.55 |
| Emphasis on Positive Multiculturalism | Mean | 3.73 | 3.81 | 3.72 | 3.77 | 3.64 | 3.68 | 3.65 | 3.73 |
| | Std. Dev. | 0.20 | 0.22 | 0.24 | 0.17 | 0.30 | 0.32 | 0.34 | 0.28 |
| Quality of School Rules | Mean | 3.25 | 3.26 | 3.39 | 3.43 | 3.11 | 3.10 | 3.27 | 3.24 |
| | Std. Dev. | 0.33 | 0.41 | 0.24 | 0.27 | 0.21 | 0.42 | 0.32 | 0.52 |
| Interpersonal Social Skills of Students | Mean | 3.02 | 3.09 | 3.18 | 3.19 | 2.83 | 2.91 | 3.07 | 3.11 |
| | Std. Dev. | 0.26 | 0.26 | 0.26 | 0.24 | 0.25 | 0.33 | 0.32 | 0.39 |
| School-Related Behavior of Students | Mean | 3.85 | 3.76 | 3.77 | 3.72 | 3.72 | 3.65 | 3.65 | 3.63 |
| | Std. Dev. | 0.44 | 0.15 | 0.16 | 0.19 | 0.35 | 0.09 | 0.15 | 0.20 |
| School Academic Climate | | | | | | | | | |
| Teachers' Extra Effort on Behalf of Students ^a | Mean | 3.50 | 3.60 | 3.51 | 3.61 | 3.47 | 3.51 | 3.64 | 3.66 |
| | Std. Dev. | 0.14 | 0.17 | 0.17 | 0.15 | 0.16 | 0.25 | 0.13 | 0.19 |
| Response to School Problems ^a | Mean | 2.97 | 2.97 | 2.96 | 3.01 | 2.92 | 2.96 | 3.00 | 3.02 |
| | Std. Dev. | 0.15 | 0.10 | 0.05 | 0.06 | 0.07 | 0.12 | 0.13 | 0.08 |
| Teachers Feel Efficacious ^a | Mean | 3.67 | 3.72 | 3.81 | 3.85 | 3.54 | 3.60 | 3.76 | 3.80 |
| | Std. Dev. | 0.17 | 0.16 | 0.12 | 0.12 | 0.20 | 0.25 | 0.19 | 0.26 |
| Academic Motivation Attributed to Students | Mean | 2.80 | 2.89 | 2.97 | 3.09 | 2.68 | 2.72 | 2.91 | 3.02 |
| | Std. Dev. | 0.23 | 0.23 | 0.15 | 0.18 | 0.22 | 0.27 | 0.25 | 0.34 |
| Future Academic Expectations for Students | Mean | 3.24 | 3.28 | 3.37 | 3.41 | 3.14 | 3.20 | 3.31 | 3.38 |
| | Std. Dev. | 0.22 | 0.16 | 0.14 | 0.17 | 0.29 | 0.26 | 0.28 | 0.32 |
| Parental Involvement | | | | | | | | | |
| Parent Involvement in General School Activities ^a | Mean | 2.06 | 2.09 | 2.17 | 2.21 | 2.06 | 2.02 | 2.15 | 2.16 |
| | Std. Dev. | 0.15 | 0.13 | 0.08 | 0.08 | 0.07 | 0.11 | 0.19 | 0.18 |
| Parent Valuing of Education | Mean | 2.77 | 2.84 | 2.93 | 3.03 | 2.69 | 2.75 | 2.89 | 2.94 |
| | Std. Dev. | 0.22 | 0.19 | 0.21 | 0.22 | 0.27 | 0.32 | 0.30 | 0.37 |
| Interaction with Parents about their Children ^a | Mean | 3.09 | 3.06 | 3.07 | 3.21 | 3.01 | 3.05 | 3.29 | 3.18 |
| | Std. Dev. | 0.25 | 0.19 | 0.18 | 0.19 | 0.13 | 0.20 | 0.15 | 0.17 |

^a Only Teachers Responded to the Items in These Measures

Table 8: Student Perceptions of School Climate: Comer Effect for Means and Slopes in School and Individual Level Analyses

| Measures | School Level Analysis ^a | | | | Individual Level Analysis | | | |
|---|--|------|----------------------|------|--|------|----------------------|------|
| | Differences in Mean Level ^b | | Differences in Slope | | Differences in Mean Level ^b | | Differences in Slope | |
| | b | se | b | se | b | se | b | se |
| SOCIAL CLIMATE | | | | | | | | |
| Social Relationships with Adults in School | | | | | | | | |
| Perceived School Staff Respect for Students | .07 | .042 | .01* | .019 | .06 | .032 | .04* | .016 |
| Perceived Caring of School Staff | .04 | .065 | .01 | .023 | .03 | .037 | .00 | .014 |
| Social Climate Among Students | | | | | | | | |
| Social Skills of Students | .11* | .044 | .03 | .026 | .11** | .036 | .02 | .014 |
| Use of Positive Problem Solving Strategies | .04 | .050 | -.02 | .019 | .05 | .041 | -.02 | .017 |
| Use of Negative Problem Solving Strategies (R) | .06 | .043 | -.01 | .022 | .05 | .029 | .02 | .013 |
| Perception of Physical Safety | .01 | .064 | .01 | .019 | .01 | .039 | .01 | .017 |
| Pride in the School | .08 | .074 | -.01 | .021 | .08* | .037 | -.01 | .015 |
| Attachment to the School | .15+ | .077 | -.01 | .035 | .13** | .047 | -.01 | .019 |
| Comfort in Interpersonal Contexts | .06 | .054 | .00 | .019 | .06 | .038 | .01 | .015 |
| Personal Feelings of Not Belonging in School (R) | -.02 | .056 | .00 | .016 | .00 | .034 | -.01 | .015 |
| ACADEMIC CLIMATE | | | | | | | | |
| Academic Relationships with Adults in School | | | | | | | | |
| Teachers' Encouragement of Academic Success | .02 | .057 | .01 | .018 | .03 | .036 | .01 | .015 |
| Teachers' Concern with Childrens' Learning | .07 | .052 | .04* | .016 | .07** | .028 | .04*** | .012 |
| Behaviors to Motivate Students | .10 | .062 | .03+ | .018 | .10*** | .029 | .03* | .011 |
| Recognition for Achievement | .11 | .117 | .06 | .036 | .08 | .055 | .05* | .023 |
| Academic Values Among Students | | | | | | | | |
| Student Devaluing of Academic Success (R) | .12+ | .059 | .02 | .021 | .12*** | .033 | .02 | .014 |
| Student Acceptance of School Values | .12* | .050 | .03 | .023 | .11*** | .026 | .02* | .011 |
| Personal Valuing of Education | .08 | .056 | .02 | .017 | .06 | .033 | .01 | .012 |

+ = p<.10 * = p<.05 ** = p<.01 *** = p<.001

^a Due to differences in statistical power between the school- and individual-level analyses, p<.10 is only reported at the school-level

^b Estimated at the final time point

(R) = Reverse Coded

Table 9: Student Climate Means by Grade and Comer Status - School Level

| Measures | Grade | LONGITUDINAL DESIGN | | | | | | | |
|---|-----------|---------------------|------|------|------|--------------|------|------|------|
| | | CONTROL (N=9) | | | | COMER (N=10) | | | |
| | | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 |
| Social Climate | | | | | | | | | |
| Social Relationships with Adults in School | | | | | | | | | |
| Perceived School Staff Respect for Students | Mean | 3.64 | 3.60 | 3.47 | 3.49 | 3.73 | 3.63 | 3.52 | 3.54 |
| | Std. Dev. | 0.23 | 0.26 | 0.15 | 0.11 | 0.17 | 0.21 | 0.12 | 0.10 |
| Perceived Caring of School Staff | Mean | 3.73 | 3.65 | 3.56 | 3.61 | 3.80 | 3.72 | 3.61 | 3.64 |
| | Std. Dev. | 0.18 | 0.14 | 0.13 | 0.13 | 0.15 | 0.16 | 0.15 | 0.23 |
| Social Climate Among Students | | | | | | | | | |
| Social Skills of Students | Mean | 3.31 | 3.21 | 3.08 | 3.15 | 3.31 | 3.30 | 3.16 | 3.21 |
| | Std. Dev. | 0.19 | 0.11 | 0.08 | 0.13 | 0.20 | 0.11 | 0.12 | 0.10 |
| Use of Positive Conflict Resolution Strategies | Mean | 2.95 | 2.82 | 2.70 | 2.71 | 3.03 | 2.93 | 2.70 | 2.75 |
| | Std. Dev. | 0.23 | 0.18 | 0.09 | 0.17 | 0.13 | 0.19 | 0.10 | 0.17 |
| Use of Negative Conflict Resolution Strategies (R) | Mean | 2.55 | 2.50 | 2.55 | 2.64 | 2.57 | 2.58 | 2.57 | 2.68 |
| | Std. Dev. | 0.31 | 0.18 | 0.15 | 0.16 | 0.25 | 0.17 | 0.17 | 0.16 |
| Perception of Physical Safety | Mean | 3.31 | 3.58 | 3.70 | 3.82 | 3.37 | 3.50 | 3.78 | 3.87 |
| | Std. Dev. | 0.24 | 0.20 | 0.12 | 0.16 | 0.15 | 0.21 | 0.15 | 0.13 |
| Pride in the School | Mean | 3.32 | 3.22 | 3.11 | 3.12 | 3.41 | 3.36 | 3.23 | 3.21 |
| | Std. Dev. | 0.22 | 0.24 | 0.21 | 0.21 | 0.19 | 0.21 | 0.19 | 0.27 |
| Attachment to the School | Mean | 3.39 | 3.12 | 3.11 | 3.12 | 3.52 | 3.47 | 3.29 | 3.22 |
| | Std. Dev. | 0.23 | 0.21 | 0.18 | 0.19 | 0.23 | 0.17 | 0.11 | 0.27 |
| Comfort in Interpersonal Contexts | Mean | 3.53 | 3.44 | 3.45 | 3.56 | 3.48 | 3.53 | 3.55 | 3.62 |
| | Std. Dev. | 0.19 | 0.20 | 0.18 | 0.14 | 0.14 | 0.15 | 0.20 | 0.23 |
| Personal Feelings of Not Belonging in School (R) | Mean | 3.48 | 3.60 | 3.74 | 3.78 | 3.42 | 3.62 | 3.71 | 3.77 |
| | Std. Dev. | 0.15 | 0.16 | 0.11 | 0.14 | 0.12 | 0.12 | 0.19 | 0.15 |
| Academic Climate | | | | | | | | | |
| Academic Relationships with Adults in School | | | | | | | | | |
| Teachers' Encouragement of Academic Success | Mean | 4.01 | 3.98 | 3.94 | 3.94 | 4.01 | 3.99 | 3.97 | 3.95 |
| | Std. Dev. | 0.09 | 0.08 | 0.10 | 0.13 | 0.15 | 0.17 | 0.12 | 0.19 |
| Teachers' Concern with Childrens' Learning | Mean | 4.04 | 3.99 | 3.94 | 3.92 | 4.01 | 4.00 | 3.96 | 3.99 |
| | Std. Dev. | 0.12 | 0.14 | 0.12 | 0.10 | 0.17 | 0.19 | 0.20 | 0.17 |
| Teachers' Behaviors to Motivate Students | Mean | 3.76 | 3.68 | 3.60 | 3.59 | 3.75 | 3.75 | 3.67 | 3.69 |
| | Std. Dev. | 0.08 | 0.11 | 0.14 | 0.14 | 0.11 | 0.11 | 0.13 | 0.16 |
| Recognition for Achievement | Mean | 3.72 | 3.79 | 3.55 | 3.48 | 3.79 | 3.74 | 3.73 | 3.58 |
| | Std. Dev. | 0.19 | 0.20 | 0.19 | 0.25 | 0.17 | 0.26 | 0.27 | 0.39 |
| Academic Values Among Students | | | | | | | | | |
| Student Devaluing of Academic Success (R) | Mean | 3.29 | 3.34 | 3.42 | 3.48 | 3.34 | 3.33 | 3.43 | 3.54 |
| | Std. Dev. | 0.10 | 0.08 | 0.08 | 0.15 | 0.14 | 0.19 | 0.12 | 0.12 |
| Student Acceptance of School Values | Mean | 3.54 | 3.42 | 3.32 | 3.39 | 3.52 | 3.48 | 3.43 | 3.47 |
| | Std. Dev. | 0.12 | 0.06 | 0.09 | 0.15 | 0.17 | 0.11 | 0.08 | 0.15 |
| Personal Valuing of Education | Mean | 3.23 | 3.14 | 3.09 | 3.13 | 3.24 | 3.23 | 3.16 | 3.24 |
| | Std. Dev. | 0.20 | 0.11 | 0.15 | 0.12 | 0.11 | 0.09 | 0.08 | 0.17 |

(R) = Reverse Coded

Table 10: Student Outcomes: Comer Effect for Means and Slopes in School and Individual Level Analyses

| Measures | School Level Analyses ^a | | | | Individual Level Analyses | | | |
|---|--|-------|----------------------|-------|--|------|----------------------|------|
| | Differences in Mean Level ^b | | Differences in Slope | | Differences in Mean Level ^b | | Differences in Slope | |
| | b | se | b | se | b | se | b | se |
| Mental Health Attributes | | | | | | | | |
| Self-Efficacy in School | .04 | .040 | .02+ | .012 | .03 | .027 | .02* | .010 |
| Racial Pride | -.01 | .040 | .02 | .020 | -.01 | .037 | .02 | .016 |
| Positive Expectations for Future Life | .07 | .052 | .03+ | .019 | .05 | .032 | .03 | .013 |
| Valuing of Positive Outcomes in Future Life | .04 | .086 | .02 | .015 | .06 | .051 | .02 | .011 |
| Satisfaction with Life and Relationships | .01 | .047 | .02 | .014 | .01 | .030 | .01 | .012 |
| Use of Constructive Coping Strategies | .04 | .055 | .03 | .022 | .03 | .033 | .03* | .014 |
| Lack of Depression | .05 | .044 | .03 | .018 | .04 | .040 | .02 | .016 |
| Lack of Anger | .17** | .054 | .05* | .021 | .16** | .050 | .04 | .019 |
| Use of Passive Coping Strategies (R) | .04 | .039 | .02 | .014 | .05 | .029 | .02 | .013 |
| Negative Social Behaviors | | | | | | | | |
| Acting Out | -.45* | .192 | -.19* | 0.078 | -.43** | .143 | -.17** | .055 |
| Disapproval of Misbehavior | -.12 | .074 | -.06* | .026 | -.12** | .042 | -.06*** | .017 |
| Substance Use | -.04 | .087 | -.05 | .034 | -.03 | .051 | -.04* | .020 |
| Positive Social Behaviors | | | | | | | | |
| Time Spent in Clubs or Lessons | .005 | .007 | -.002 | .003 | .006 | .006 | -.001 | .003 |
| Time Spent Playing Sports | .002 | .001 | .001 | .003 | .003 | .006 | .001 | .003 |
| Time Spent Doing Homework or Free Reading | .006 | .010 | .004 | .004 | .005 | .007 | .004 | .030 |
| Time Spent in Unstructured Activities | -.017 | .012 | -.002 | .005 | -.16 | .008 | -.002 | .003 |
| Academic Achievement | | | | | | | | |
| Math NCE Test Score | .72 | 1.144 | 1.38** | .418 | .81 | .632 | 1.41*** | .229 |
| Reading NCE Test Score | .09 | 1.177 | .95** | .309 | .27 | .659 | .91*** | .256 |

+ = p<.10 * = p<.05 ** = p<.01 *** = p<.001

^a Due to differences in statistical power between the school- and individual-level analyses, p<.10 is only reported at the school-level

^b Estimated at the final time point

(R) = Reverse Coded

Table 11: Student Outcome Means by Grade and Comer Status - School Level

| Measures | Grade | CONTROL (N=9) | | | | COMER (N=10) | | | |
|---|-----------|---------------|-------|-------|-------|--------------|-------|-------|-------|
| | | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 |
| Mental Health Attributes | | | | | | | | | |
| Self-Efficacy in School | Mean | 3.76 | 3.81 | 3.74 | 3.80 | 3.71 | 3.79 | 3.81 | 3.86 |
| | Std. Dev. | 0.17 | 0.13 | 0.12 | 0.10 | 0.17 | 0.12 | 0.11 | 0.18 |
| Racial Pride | Mean | 3.98 | 4.03 | 3.96 | 4.08 | 3.91 | 3.92 | 3.97 | 4.06 |
| | Std. Dev. | 0.13 | 0.10 | 0.14 | 0.14 | 0.15 | 0.16 | 0.18 | 0.16 |
| Positive Expectations for Future Life | Mean | 4.24 | 4.22 | 4.17 | 4.09 | 4.16 | 4.25 | 4.19 | 4.20 |
| | Std. Dev. | 0.13 | 0.14 | 0.10 | 0.13 | 0.14 | 0.14 | 0.10 | 0.14 |
| Valuing of Positive Outcomes in Future Life | Mean | 3.99 | 3.95 | 3.83 | 3.79 | 3.99 | 3.97 | 3.92 | 3.90 |
| | Std. Dev. | 0.19 | 0.14 | 0.16 | 0.14 | 0.12 | 0.14 | 0.14 | 0.20 |
| Satisfaction with Life and Relationships | Mean | 3.98 | 4.01 | 4.00 | 4.06 | 3.96 | 4.02 | 3.99 | 4.12 |
| | Std. Dev. | 0.17 | 0.18 | 0.09 | 0.10 | 0.13 | 0.11 | 0.15 | 0.19 |
| Use of Constructive Coping Strategies | Mean | 4.17 | 4.22 | 4.10 | 4.05 | 4.17 | 4.19 | 4.10 | 4.11 |
| | Std. Dev. | 0.10 | 0.13 | 0.07 | 0.14 | 0.12 | 0.10 | 0.12 | 0.18 |
| Lack of Depression | Mean | 3.54 | 3.75 | 3.97 | 4.02 | 3.51 | 3.76 | 3.98 | 4.11 |
| | Std. Dev. | 0.15 | 0.14 | 0.12 | 0.12 | 0.17 | 0.10 | 0.11 | 0.14 |
| Lack of Anger | Mean | 3.02 | 3.14 | 3.22 | 3.21 | 3.07 | 3.18 | 3.29 | 3.36 |
| | Std. Dev. | 0.18 | 0.11 | 0.19 | 0.19 | 0.19 | 0.13 | 0.19 | 0.13 |
| Use of Passive Coping Strategies (R) | Mean | 3.96 | 4.11 | 4.11 | 4.26 | 3.96 | 4.04 | 4.17 | 4.30 |
| | Std. Dev. | 0.07 | 0.10 | 0.12 | 0.11 | 0.18 | 0.17 | 0.14 | 0.13 |
| Negative Social Behaviors | | | | | | | | | |
| Acting Out | Mean | 3.08 | 3.22 | 3.74 | 4.11 | 3.06 | 3.28 | 3.54 | 3.54 |
| | Std. Dev. | 0.58 | 0.38 | 0.43 | 0.35 | 0.54 | 0.65 | 0.62 | 0.56 |
| Disapproval of Misbehavior | Mean | 4.10 | 4.08 | 3.95 | 3.85 | 4.07 | 4.07 | 3.97 | 4.02 |
| | Std. Dev. | 0.20 | 0.12 | 0.18 | 0.19 | 0.16 | 0.19 | 0.20 | 0.18 |
| Substance Use | Mean | 0.30 | 0.34 | 0.57 | 0.71 | 0.45 | 0.44 | 0.58 | 0.68 |
| | Std. Dev. | 0.12 | 0.09 | 0.19 | 0.19 | 0.23 | 0.13 | 0.18 | 0.23 |
| Positive Social Behaviors | | | | | | | | | |
| Time Spent in Clubs or Lessons | Mean | 0.20 | 0.17 | 0.15 | 0.13 | 0.19 | 0.19 | 0.16 | 0.13 |
| | Std. Dev. | 0.03 | 0.03 | 0.02 | 0.02 | 0.03 | 0.02 | 0.03 | 0.03 |
| Time Spent Playing Sports | Mean | 0.19 | 0.22 | 0.23 | 0.23 | 0.20 | 0.21 | 0.24 | 0.23 |
| | Std. Dev. | 0.03 | 0.02 | 0.02 | 0.02 | 0.01 | 0.02 | 0.02 | 0.03 |
| Time Spent Doing Homework or Free Reading | Mean | 0.29 | 0.26 | 0.24 | 0.23 | 0.29 | 0.25 | 0.23 | 0.24 |
| | Std. Dev. | 0.06 | 0.02 | 0.03 | 0.02 | 0.05 | 0.03 | 0.03 | 0.03 |
| Time Spent in Unstructured Activities | Mean | 0.35 | 0.37 | 0.40 | 0.42 | 0.34 | 0.36 | 0.39 | 0.40 |
| | Std. Dev. | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.02 | 0.04 |
| Academic Achievement | | | | | | | | | |
| Math NCE Test Score | Mean | 33.72 | 34.39 | 31.68 | 36.75 | 30.38 | 30.72 | 30.33 | 36.72 |
| | Std. Dev. | 8.50 | 9.04 | 7.34 | 6.37 | 6.35 | 5.70 | 7.27 | 5.87 |
| Reading NCE Test Score | Mean | 36.78 | 34.64 | 34.97 | 37.27 | 33.60 | 33.11 | 34.43 | 37.93 |
| | Std. Dev. | 4.49 | 4.98 | 3.00 | 3.11 | 3.79 | 4.06 | 4.34 | 4.87 |

(R) = Reverse Coded

Table 12: School-Level Means for Achievement, Acting Out, and Disapproval of Misbehavior by Principal Change

| | | | | 5 | 6 | 7 | 8 |
|-----------------------------------|-----------|-----------|-----------|-------|-------|-------|-------|
| Math | No Change | Control | Mean | 35.90 | 37.59 | 32.51 | 37.19 |
| | | (N=4) | Std. Dev. | 3.92 | 7.54 | 4.88 | 4.35 |
| | | Comer | Mean | 29.81 | 30.77 | 30.73 | 36.16 |
| | | (N=9) | Std. Dev. | 6.47 | 6.04 | 7.59 | 5.94 |
| | Change | Control | Mean | 31.98 | 31.83 | 31.01 | 36.40 |
| | | (N=5) | Std. Dev. | 11.15 | 10.12 | 9.41 | 8.16 |
| | Comer | Mean | 35.48 | 30.23 | 26.74 | 41.73 | |
| | (N=1) | Std. Dev. | . | . | . | . | |
| Reading | No Change | Control | Mean | 38.42 | 36.75 | 36.82 | 38.23 |
| | | (N=4) | Std. Dev. | 2.79 | 4.58 | 3.52 | 4.16 |
| | | Comer | Mean | 33.56 | 33.18 | 34.70 | 37.33 |
| | | (N=9) | Std. Dev. | 4.01 | 4.30 | 4.52 | 4.77 |
| | Change | Control | Mean | 35.46 | 32.96 | 33.50 | 36.51 |
| | | (N=5) | Std. Dev. | 5.44 | 5.10 | 1.61 | 2.17 |
| | Comer | Mean | 33.89 | 32.51 | 32.03 | 43.22 | |
| | (N=1) | Std. Dev. | . | . | . | . | |
| Acting Out | No Change | Control | Mean | 3.03 | 2.96 | 3.90 | 4.11 |
| | | (N=4) | Std. Dev. | 0.32 | 0.13 | 0.42 | 0.39 |
| | | Comer | Mean | 3.02 | 3.29 | 3.41 | 3.41 |
| | | (N=9) | Std. Dev. | 0.55 | 0.63 | 0.48 | 0.40 |
| | Change | Control | Mean | 3.11 | 3.42 | 3.61 | 4.11 |
| | | (N=5) | Std. Dev. | 0.76 | 0.39 | 0.43 | 0.35 |
| | Comer | Mean | 3.43 | 4.06 | 4.74 | 4.72 | |
| | (N=1) | Std. Dev. | . | . | . | . | |
| Disapproval of Misbehavior | No Change | Control | Mean | 4.15 | 4.17 | 3.86 | 3.80 |
| | | (N=4) | Std. Dev. | 0.20 | 0.13 | 0.24 | 0.22 |
| | | Comer | Mean | 4.06 | 4.10 | 4.01 | 4.07 |
| | | (N=9) | Std. Dev. | 0.21 | 0.18 | 0.17 | 0.11 |
| | Change | Control | Mean | 4.06 | 4.01 | 4.03 | 3.89 |
| | | (N=5) | Std. Dev. | 0.17 | 0.05 | 0.10 | 0.17 |
| | Comer | Mean | 4.16 | 3.83 | 3.64 | 3.61 | |
| | (N=1) | Std. Dev. | . | . | . | . | |

Table 13: Correlations between Outcome Changes for Grades 6 thru 8 at the School and Individual Levels - Comer Students Only^a

| | | 1 | 2 | 3 | 4 | 5 |
|---|----------------------------|--------|-------|--------|-------|-------|
| 1 | Change in Acting Out | ----- | -.46 | -.86** | .31 | .21 |
| 2 | Change in Lack of Anger | -.30** | ----- | .38 | -.16 | -.46 |
| 3 | Change in Disapproval of M | -.43** | .38** | ----- | -.43 | -.39 |
| 4 | Change in Math | -.07 | .07 | .08 | ----- | .75* |
| 5 | Change in Reading | .03 | -.04 | .02 | .24** | ----- |

* p<.05 ** p < .01

^a Values below the diagonal are computed at the individual level and values above the diagonal are computed at the school level.

Figure 1: Implementation Index by Comer Status, 1994-1997

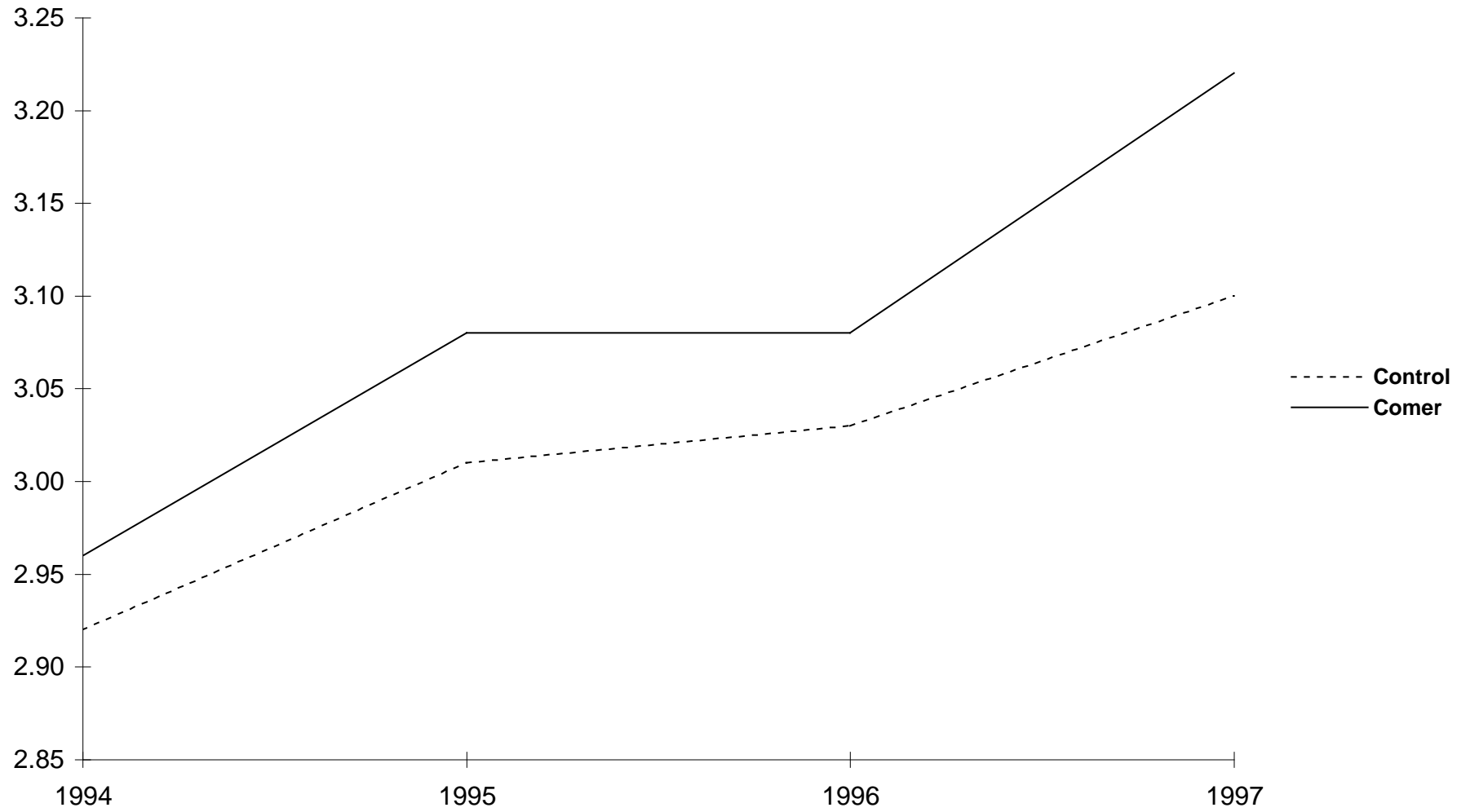


Figure 2: Staff Social Climate by Year and Comer Status

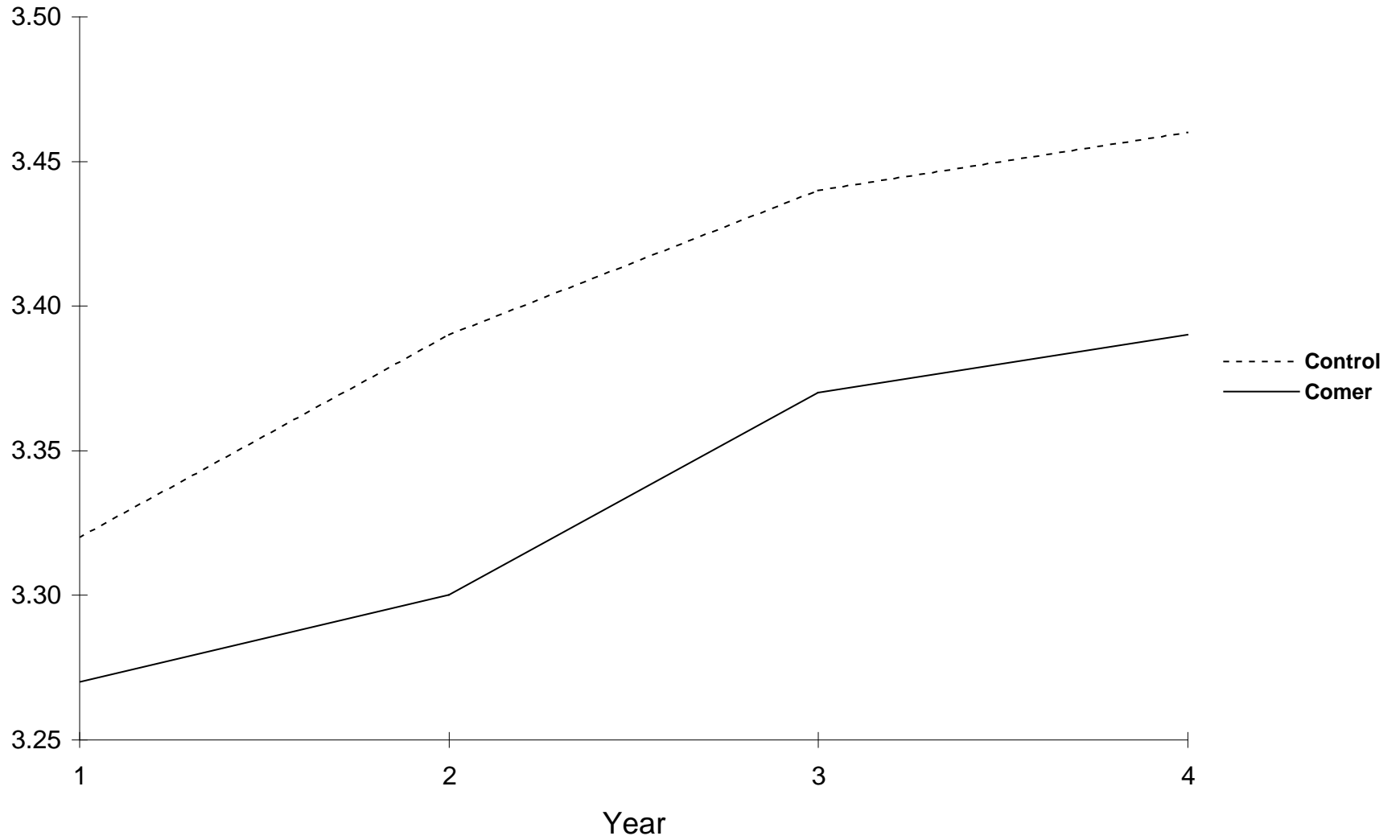


Figure 3: Student Reports of Social Climate by Grade and Comer Status -- Longitudinal Design

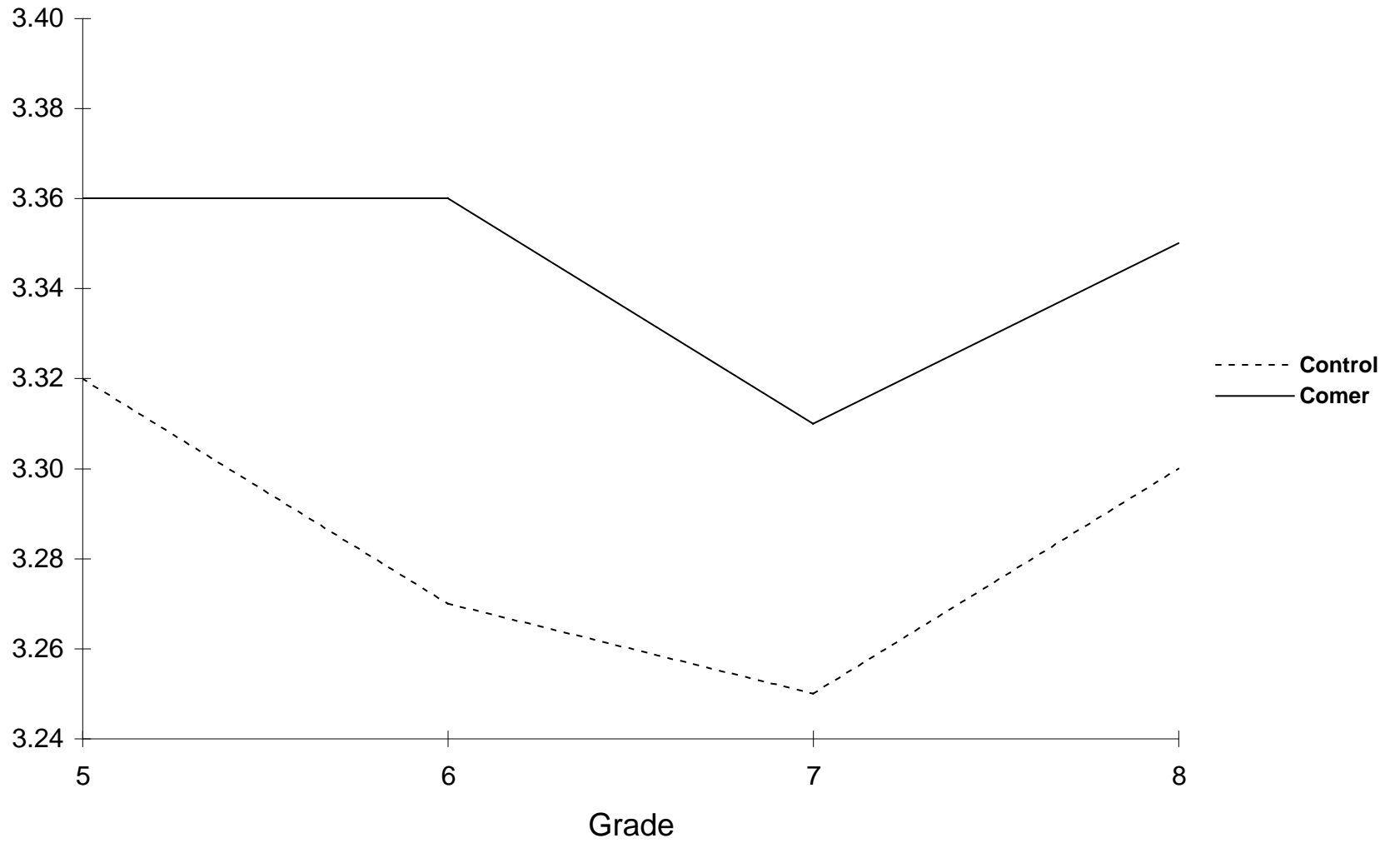


Figure 4: Staff Academic Climate by Year and Comer Status

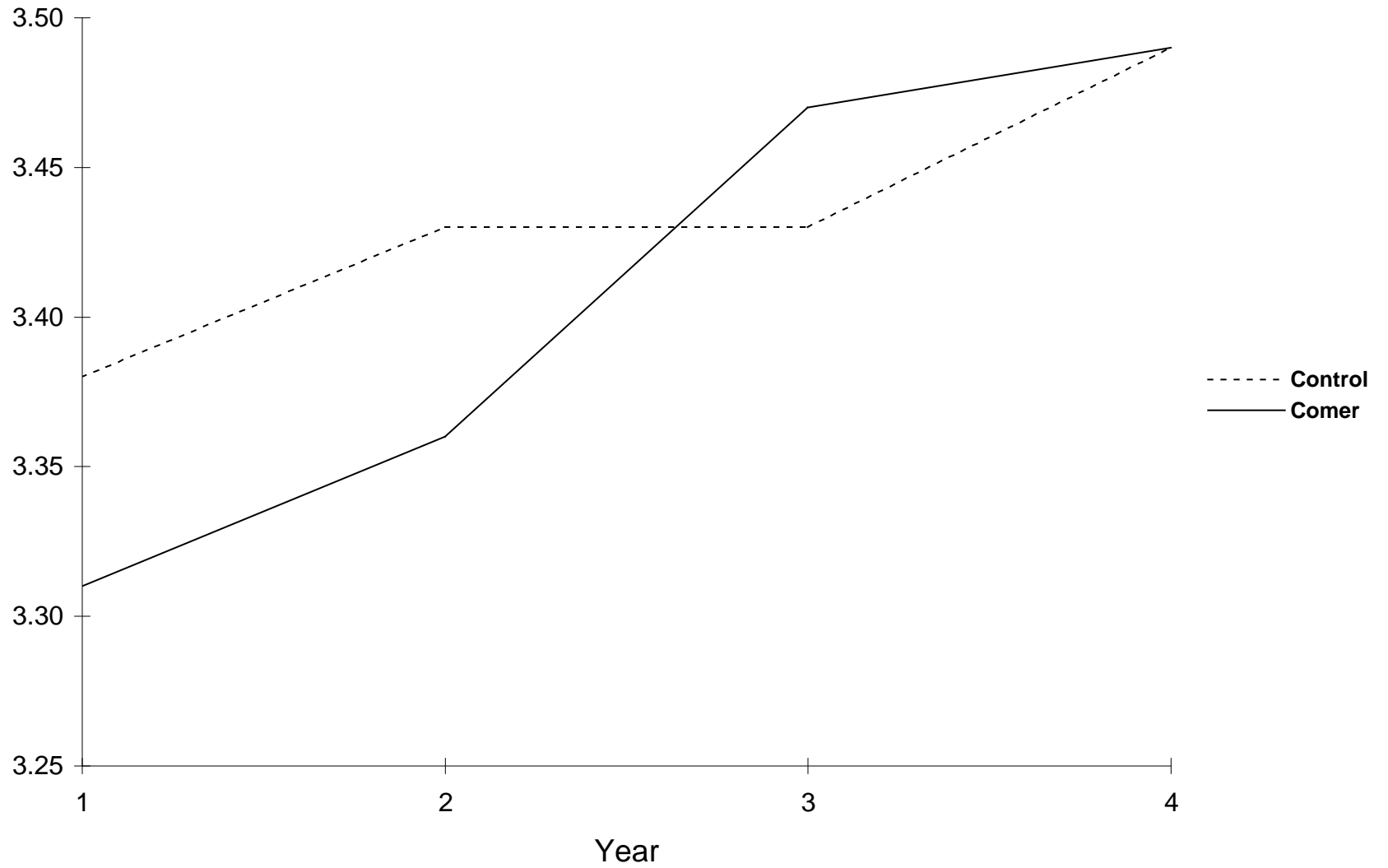


Figure 5: Student Reports of Academic Climate by Grade and Comer Status -- Longitudinal Design

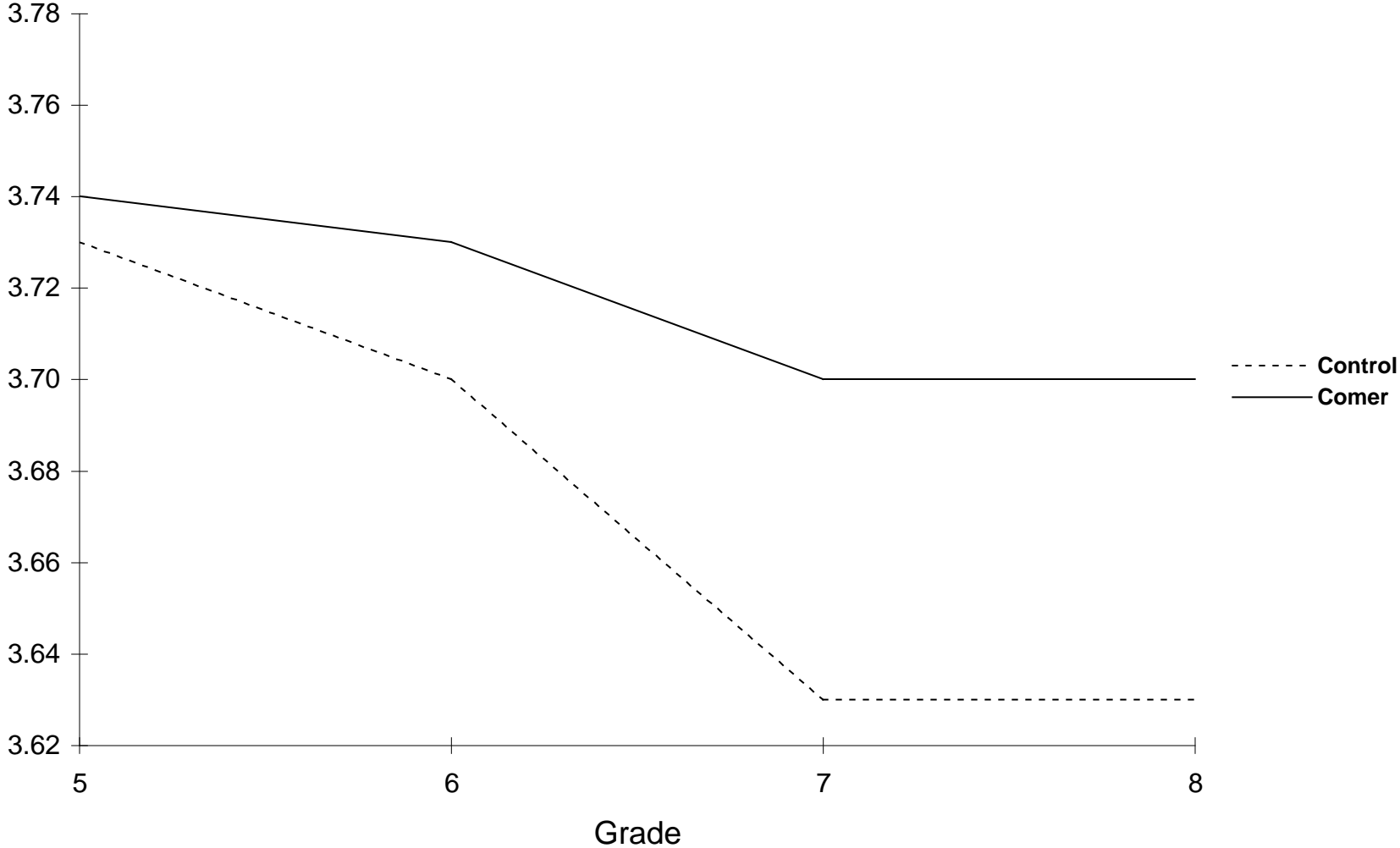


Figure 6: School-Level Means For Math Scores by Grade and Comer Status -- Longitudinal Design

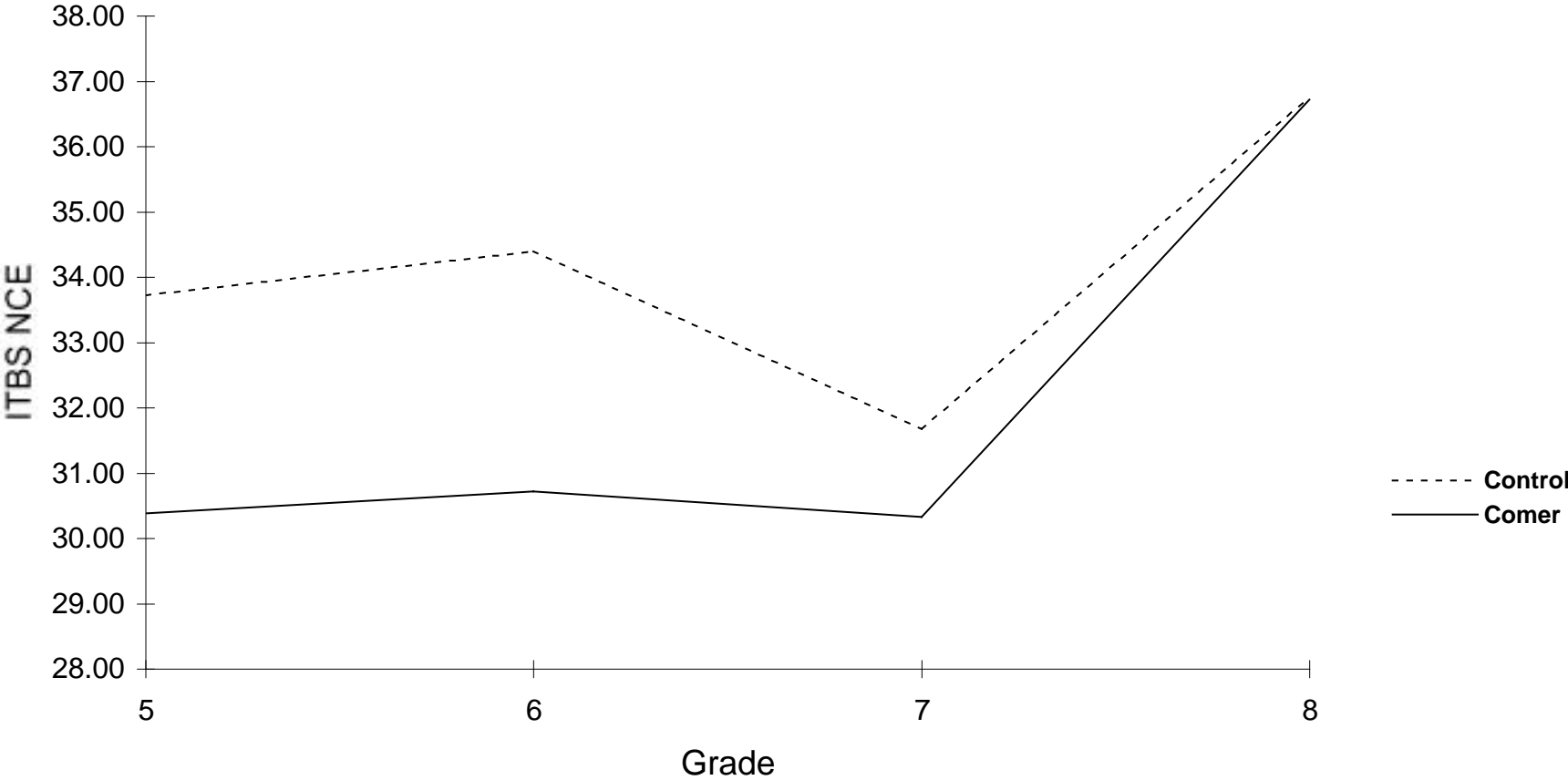


Figure 7: School Level-Means for Reading Scores by Grade and Comer Status -- Longitudinal Design

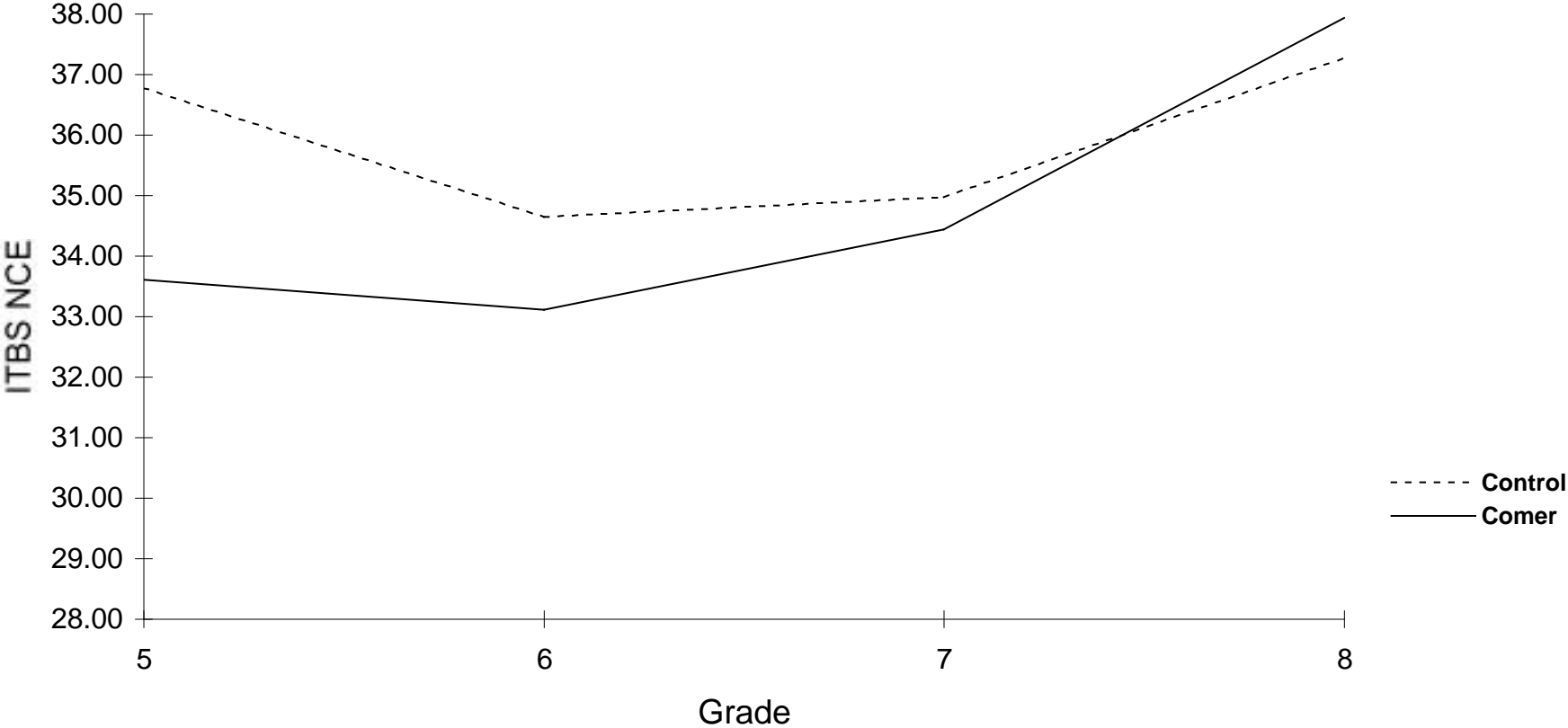


Figure 8: School-Level Means for Acting Out by Grade and Comer Status -- Longitudinal Design

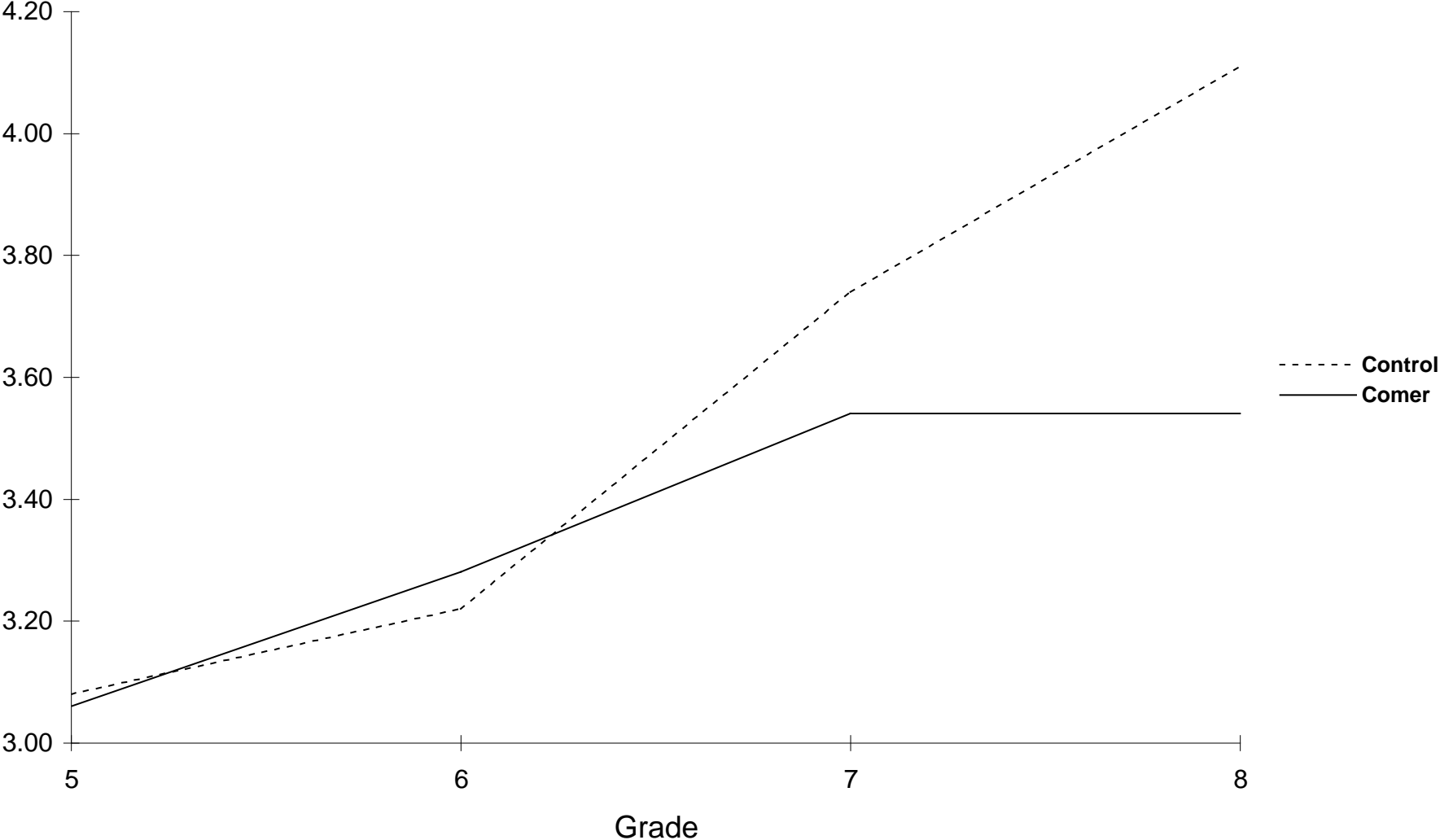


Figure 9: School-Level Means for Disapproval of Misbehavior by Grade and Comer Status -- Longitudinal Design

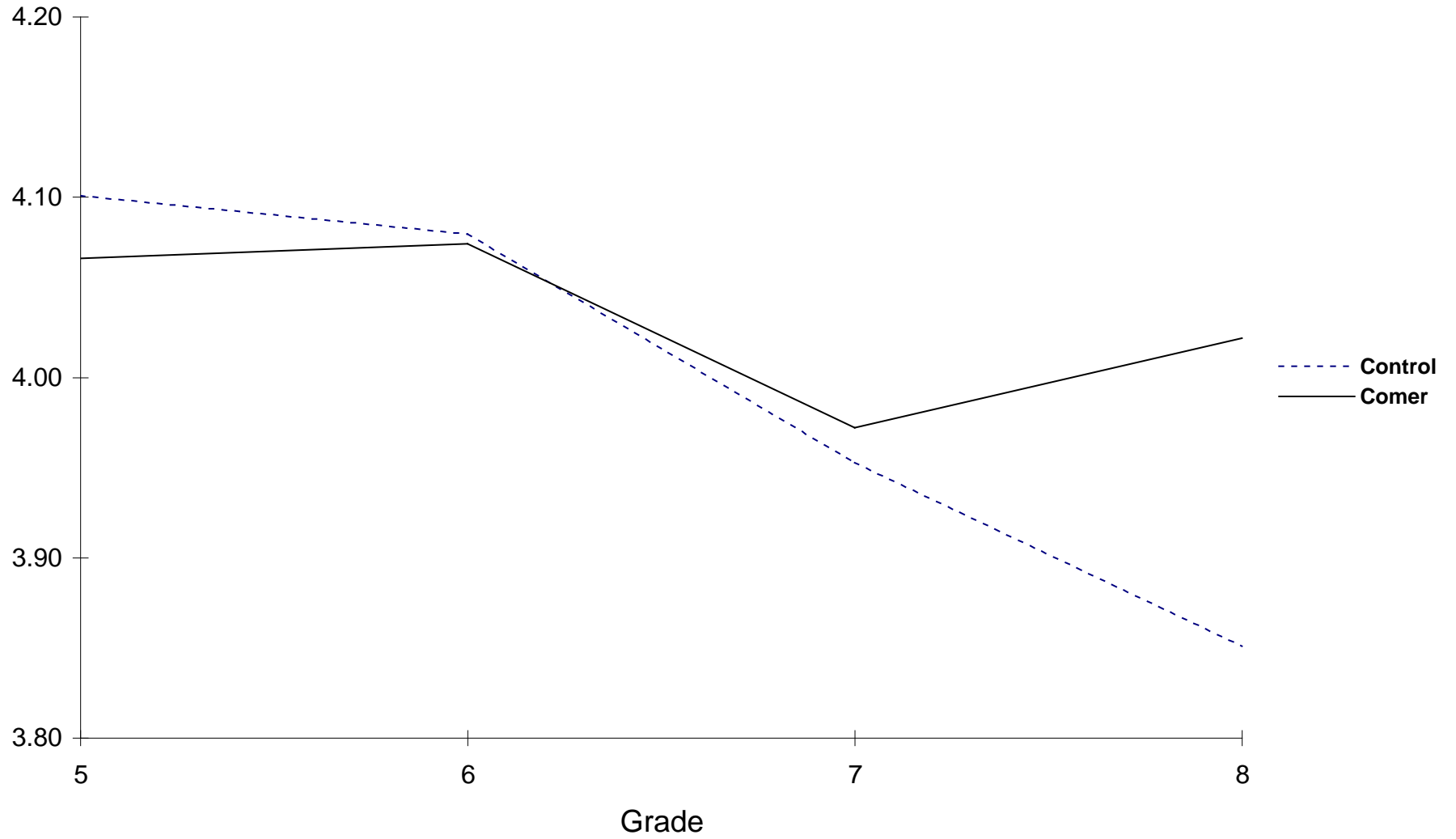


Figure 10: School-Level Means for Anger Control by Grade and Comer Status -- Longitudinal Design

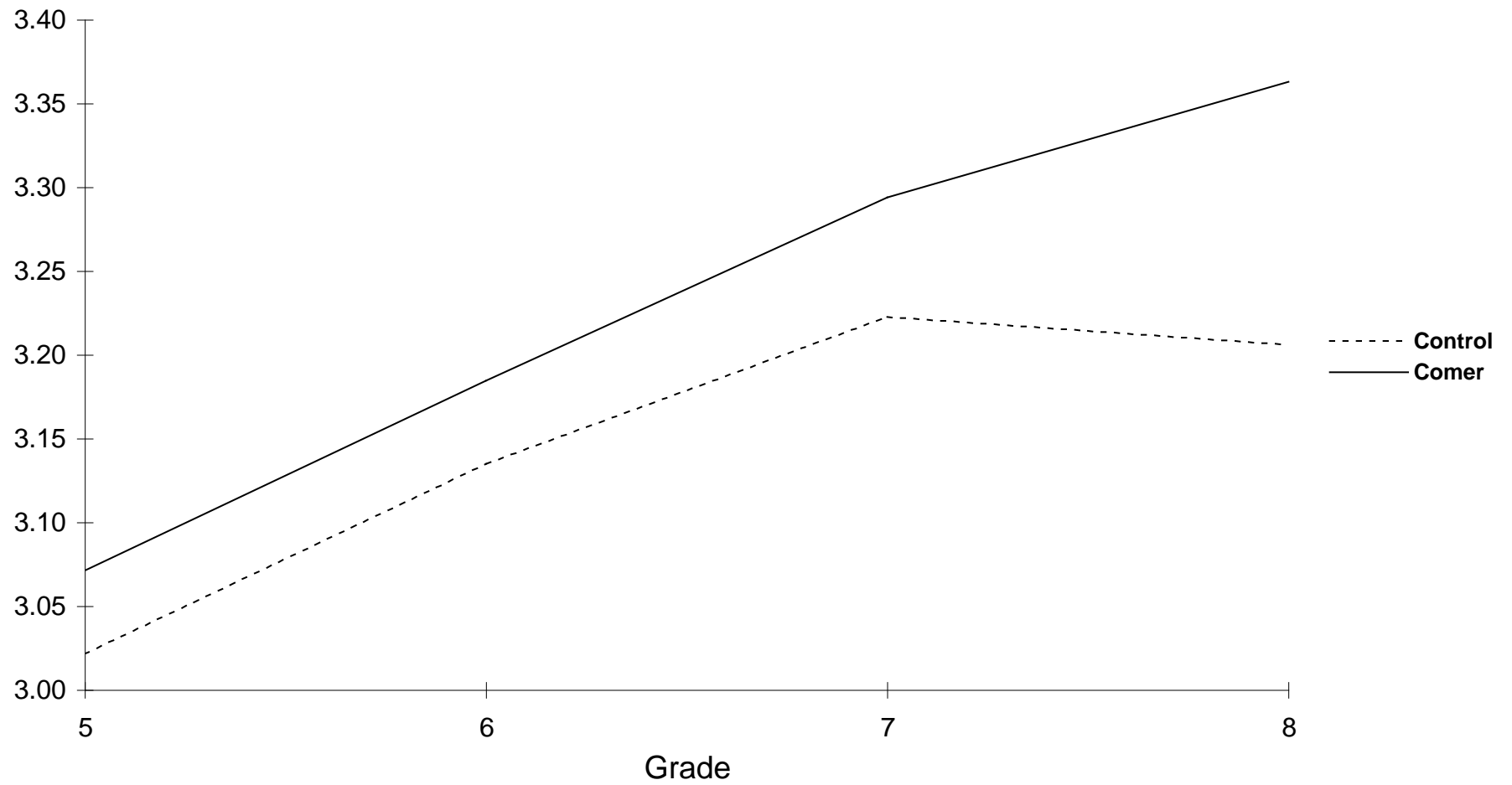


Figure 11: Student Reports of Social Climate by Year and Comer Status -- Cross-Sectional Design

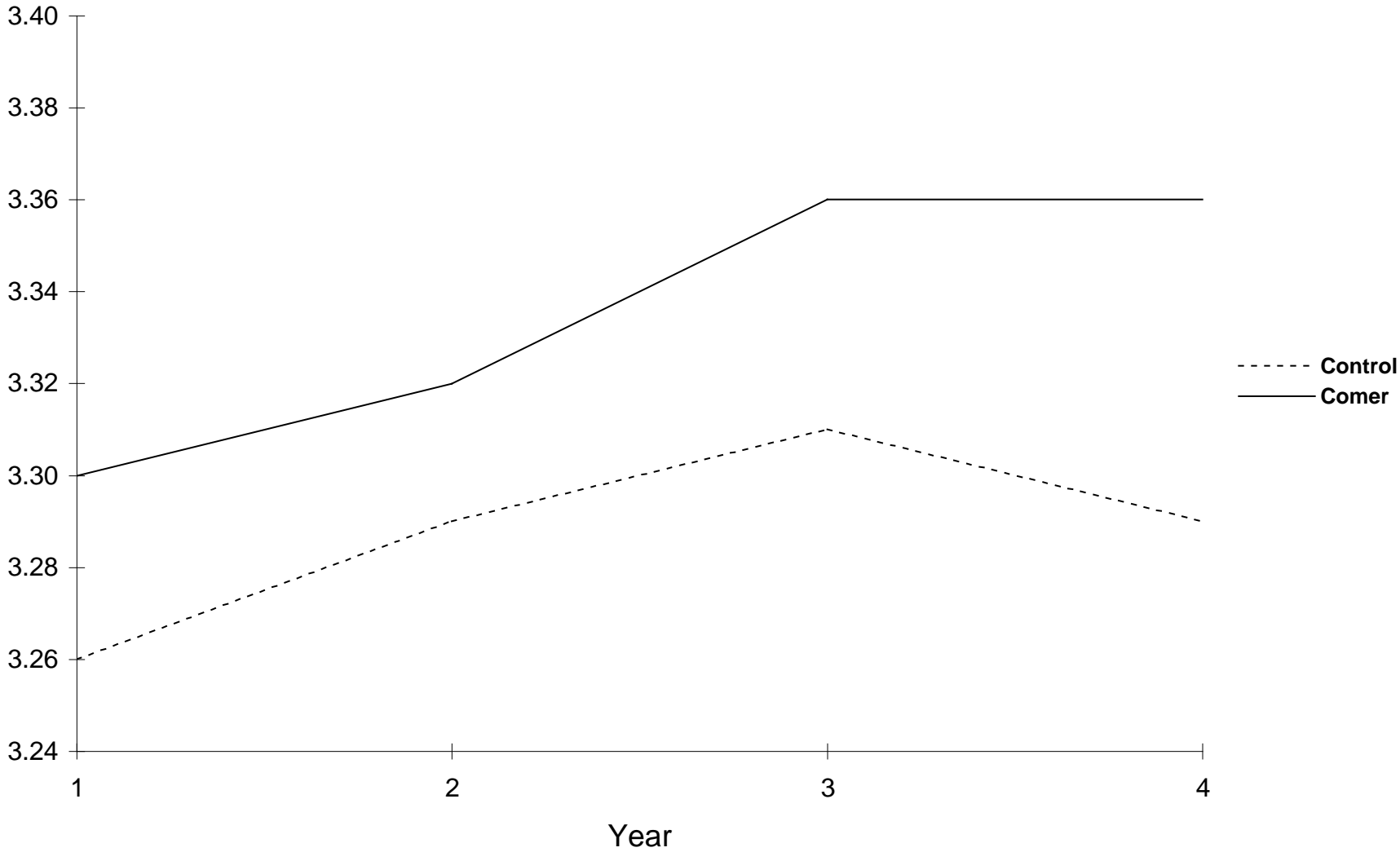


Figure 12: Student Reports of Academic Climate by Year and Comer Status -- Cross-Sectional Design

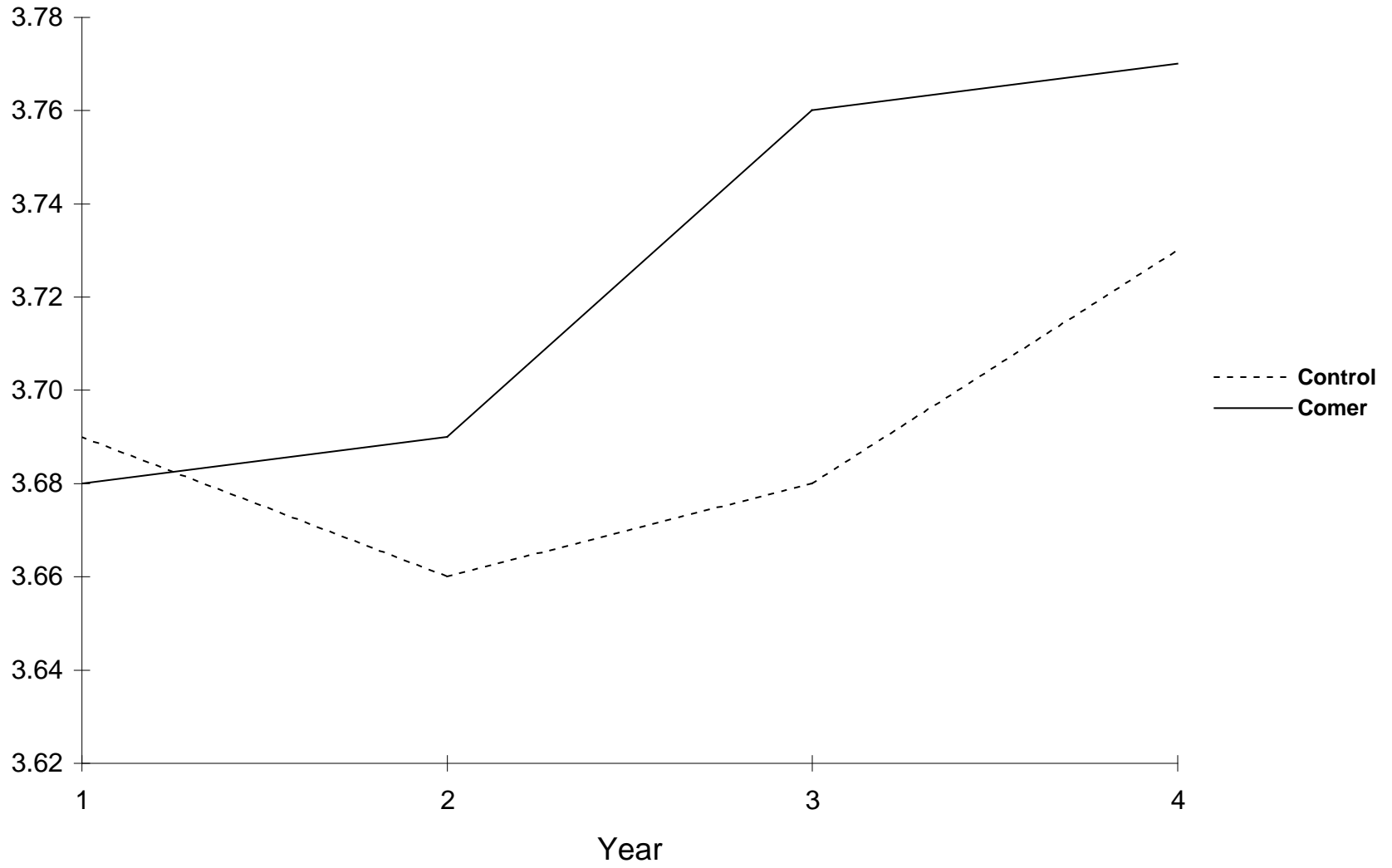


Figure 13: School-Level Means for Acting Out by Year and Comer Status -- Cross-Sectional Design

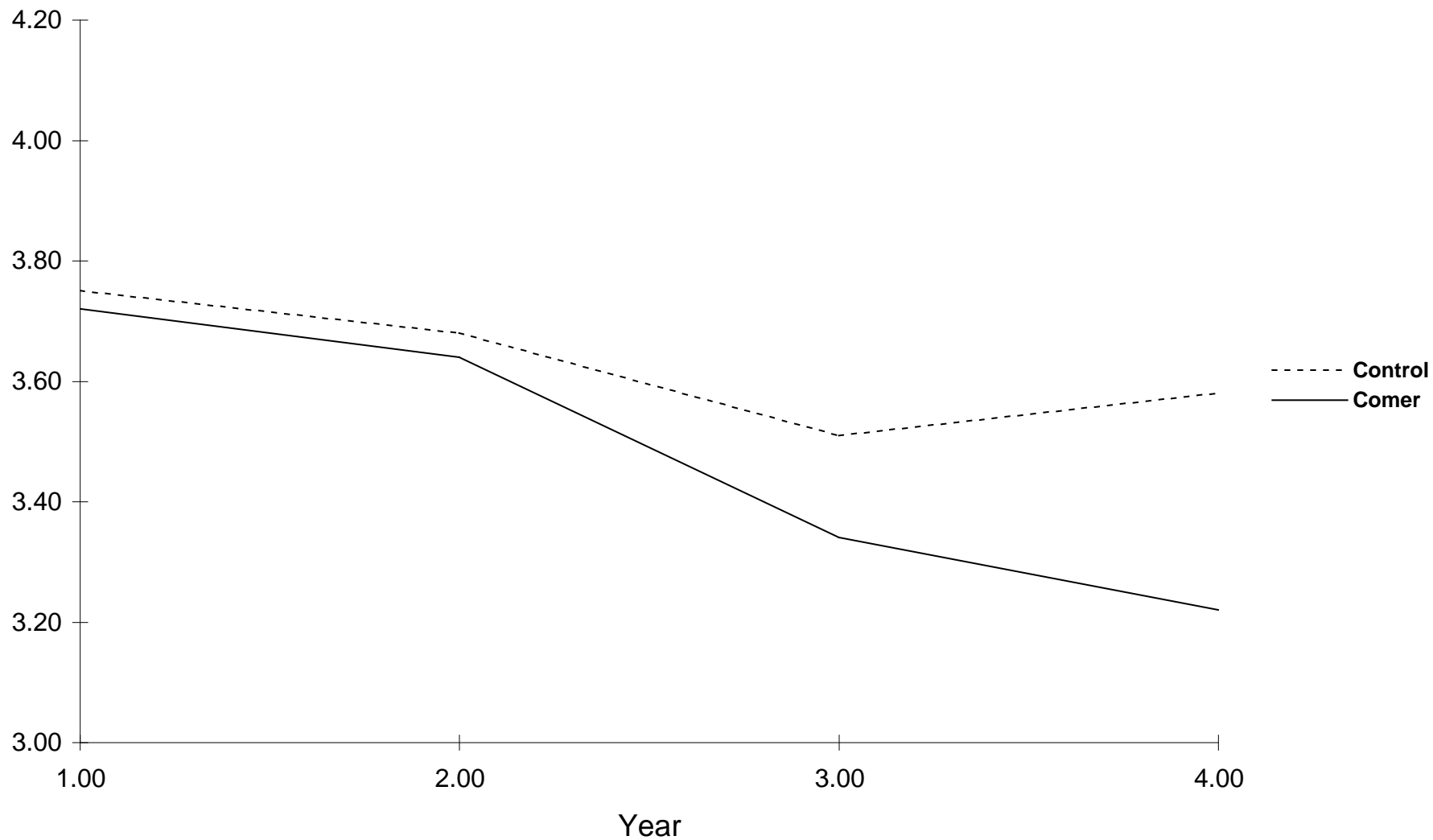


Figure 14: School-Level Means for Disapproval of Misbehavior by Year by Comer Status --
Cross-Sectional Design

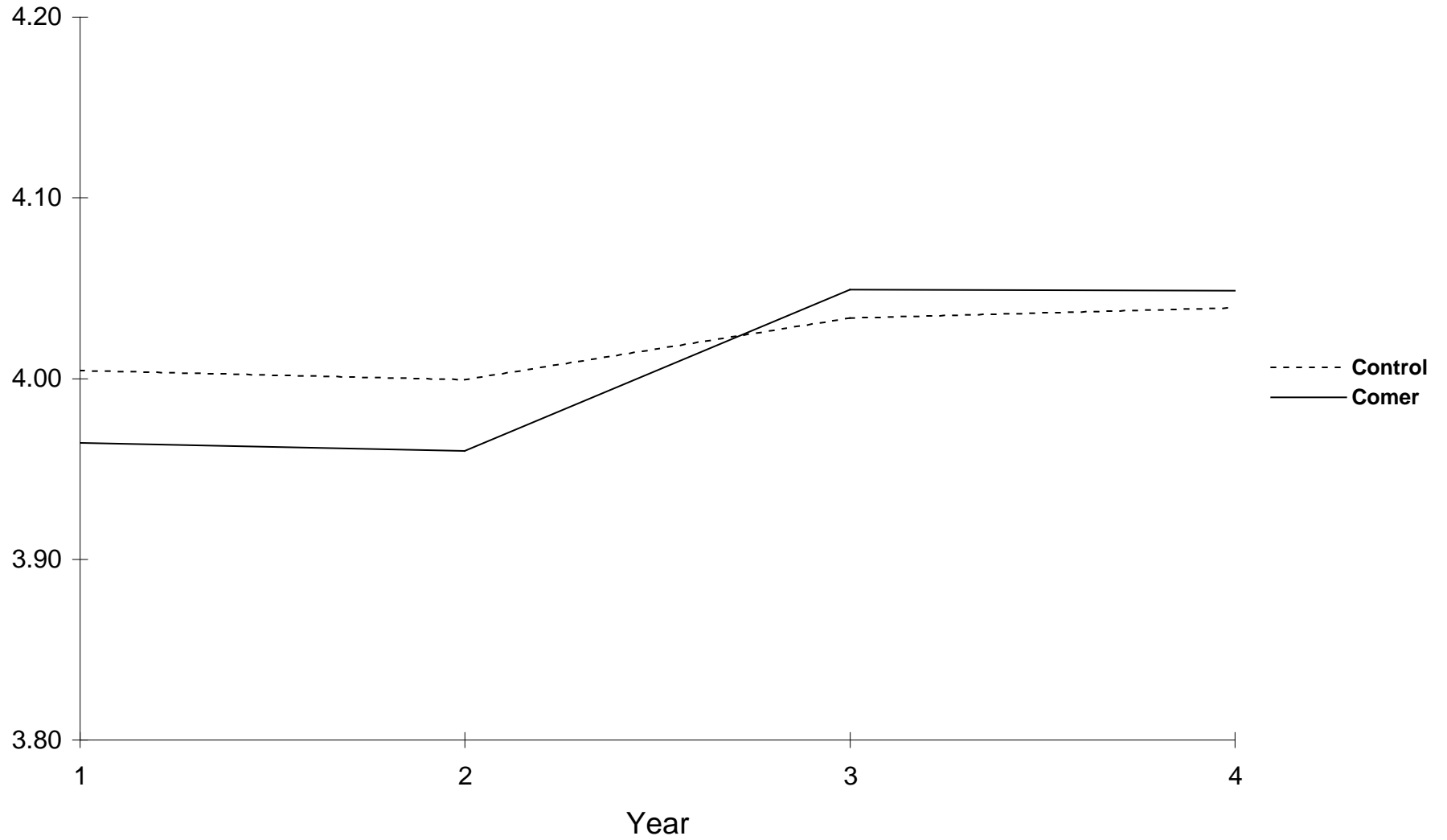


Figure 15: School-Level Means for Anger Control by Year and Comer Status -- Cross-Sectional Design

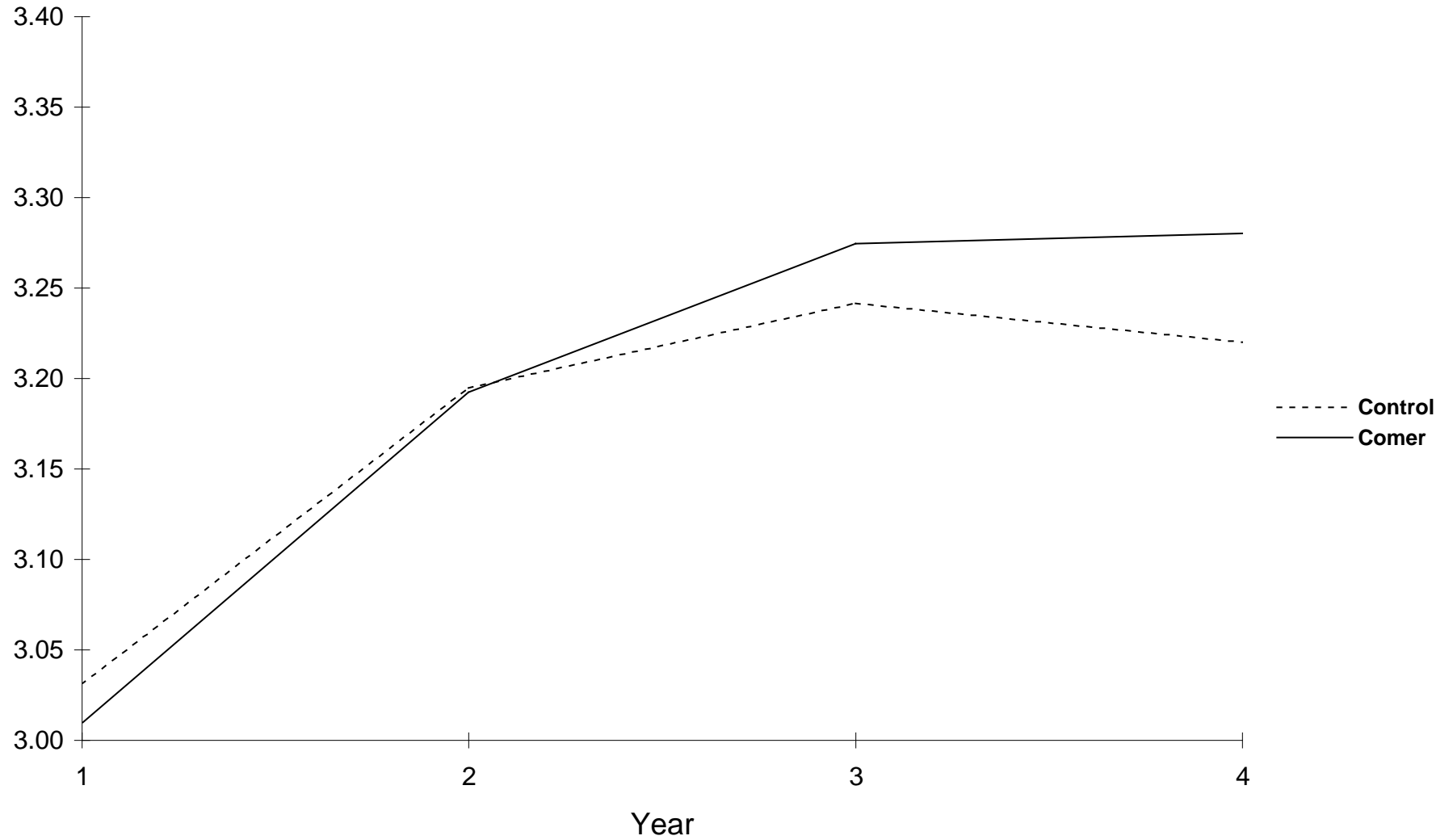


Figure 16: School-Level Means for Math Scores by Year and Comer Status -- Cross-Sectional Design

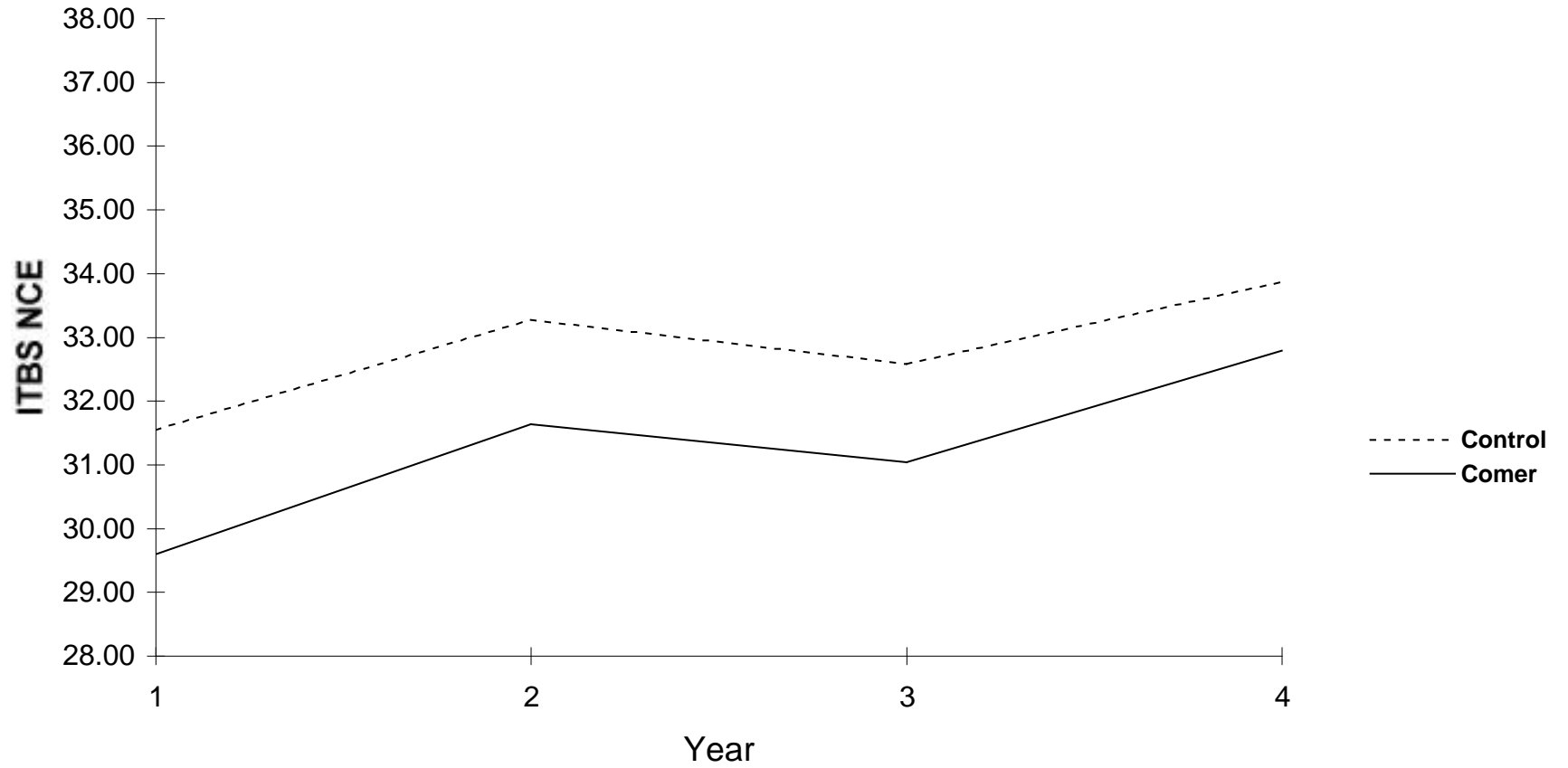


Figure 17: School-Level Means for Reading Scores by Year and Comer Status -- Cross-Sectional Design

