

After-School Programs for High School Students

An Evaluation of After School Matters

Technical Report

Barton J. Hirsch

Larry V. Hedges

Northwestern University

JulieAnn Stawicki

University of Wisconsin – Extension

Megan A. Mekinda

Northwestern University



Cover photo courtesy of Gloria Hafer

June 2011

© 2011 by B. Hirsch, L. Hedges, J. Stawicki, & M. Mekinda

TABLE OF CONTENTS

Lists of Exhibits	ii
Acknowledgements	iii
Executive Summary	1
Chapter 1 Overview	8
Background	8
After School Matters	13
Research Questions	18
Chapter 2 Research Methods	24
Sample Selection.....	24
Participants.....	28
Data Collection/Research Procedures	31
Measures	32
Data Analysis Overview	38
Chapter 3 Implementation	41
High/Scope.....	41
Program Design Features	42
Social Climate	43
Implementation and Outcome Measure Correlations	43
Qualitative Observations.....	44
Participation in Alternative Activities	46
Chapter 4 Impact Findings: Intent-to-Treat	47
Chapter 5 Impact Findings: Treatment-on-the-Treated	54
Chapter 6 Discussion	59
Conclusions and Future Directions	66
Appendix A: Design Features Instrument	73
References	74
Tables	83

LIST OF EXHIBITS

Tables

- 2.1 Data Collected for the Evaluation
- 2.3 Treatment Attrition
- 2.4 Study Participation Rates
- 2.5 Demographics: Entire Sample
- 2.6 Demographics: Treatment-on-the-Treated
- 2.7 Treatment-on-the-Treated/Dropout Pretest Demographics
- 2.8 Control Group Extracurricular Activities

- 3.1 Implementation Measure: High/Scope
- 3.2 Implementation Measure: Design Features
- 3.3 Implementation Measure: Social Climate
- 3.4 Implementation and Outcome Measure Correlations

- 4.1 Positive Youth Development Outcomes: Intent-to-Treat
- 4.2 Marketable Job Skills Outcomes: Intent-to-Treat
- 4.3 Academic Performance Outcomes: Intent-to-Treat
- 4.4 School Attendance with Prior Year Attendance Control: Intent-to-Treat
- 4.5 School Attendance Entire Sample: Intent-to-Treat
- 4.6 Attitudes about School: Intent-to-Treat
- 4.7 Problem Behavior Outcomes: Intent-to-Treat

- 5.1 Positive Youth Development Outcomes: Treatment-on-the-Treated
- 5.2 Marketable Job Skills Outcomes: Treatment-on-the-Treated
- 5.3 Academic Performance Outcomes: Treatment-on-the-Treated
- 5.4 School Attendance with Prior Year Attendance Control: Treatment-on-the-Treated
- 5.5 School Attendance Entire Sample: Treatment-on-the-Treated
- 5.6 Attitudes about School: Treatment-on-the-Treated
- 5.7 Problem Behavior Outcomes: Treatment-on-the-Treated

- 6.1 Average Effect Size by Outcome Domain

ACKNOWLEDGEMENTS

This report was a result of the efforts of many people who have contributed and collaborated with us during the evaluation.

We are grateful to all the staff of After School Matters, including David Sinski, Executive Director, and Rachel Klein, Raymond Legler, Natasha Smith, and Myrna Torres. We appreciate the support of Maggie Daley, Chair of the Board of Directors. Thanks to all the Regional Directors and Specialists who worked hard and assisted in recruitment and planning, and to the apprenticeship instructors who generously allowed access to their programs. Of course the study would not have been possible without the many young people who graciously participated in the research.

The study has benefited from the cooperation of Chicago Public Schools principals, staff, and teachers.

We could not have completed this study without the help of Kathadin Cook, Jaime Platzer, Deborah Puntteney, Christy Serrano, and Oseela Thomas who observed apprenticeships and conducted interviews with participants and after-school instructional staff. Special thanks to Deborah Puntteney who helped get the study off the ground in the first year.

We thank the many human resource interviewers who volunteered their time to conduct mock job interviews in the three years of data collection. We are especially grateful to Cheryl Berrington and Wilbert Williams whose contributions were invaluable in the design of the Mock Job Interview.

The study's scientific advisory board provided valuable insights on the evaluation's design and data analysis and in the preparation of this report. We thank Jacquelynne Eccles, Greg Duncan, Denise Gottfredson, Robert Halpern, Stephen Hamilton, Reed Larson, Jeylan Mortimer, and Elizabeth Reisner for their expertise and guidance. Spyros Konstantopoulos provided statistical and methodological guidance in the earliest stages of the research. Joseph Durlak helped us think through some of the issues raised in the discussion section.

Finally, we thank the William T. Grant Foundation, Wallace Foundation, and Searle Fund for funding this evaluation. We also thank After School Matters for providing a planning grant in the development stage of the study.

EXECUTIVE SUMMARY

After-school programs have attracted a strong and growing constituency among both academic theorists and policy makers. Yet participation in after-school programs is extremely sparse at the high school level. Clearly there is a need to develop more attractive after-school programs for HS youth, which makes After School Matters (ASM), one of the country's largest programs for this group, an important program to study.

After School Matters offers paid apprenticeship-type experiences in a wide array of areas, such as technology, arts, and sports. Each apprenticeship involves work in the designated area, learning and making use of relevant skills to accomplish a task. Instructors are present to provide information, guidance, and feedback, and to introduce students to the standards, language, and culture of that line of work. The experience presumably also helps students begin to appreciate and adapt to the culture of the workplace and improve the “soft skills” increasingly demanded by employers. The instructors have expertise in – and in many instances earn their livelihood through – the activity that is the focus of the apprenticeship. Most instructors are not teachers. Apprentices were paid a stipend equivalent to \$5/hour during our study.

After-school programs that have an apprenticeship orientation, such as ASM, have the potential to provide the benefits of successful part-time work experience at a lower cost than many workforce development programs. Moreover, as an after-school program, they have the latitude to focus more broadly on positive youth development than might be the case with programs targeted exclusively at workforce development. Prior research on ASM suggested that their apprenticeships could provide such an environment.

This report presents results from a three-year, random assignment evaluation of After School Matters. The major questions addressed by this evaluation are whether assignment to ASM apprenticeships results in improvements in positive youth development, marketable job skills, academic performance, and problem behavior.

Methods

The evaluation involved a randomized controlled trial. Randomization is the best procedure for guarding against selection effects in which youth more likely to improve over time are disproportionately located in the treatment group. In that event, it is impossible to sort out whether effects are due to selection (who got into the program) or the program itself (the experiences of youth while in the program). This is why randomized controlled trials are typically considered the “gold standard” in evaluation research.

We assessed 13 apprenticeships and their respective control groups for a total of 535 youth. The selection of apprenticeships focused on those instructors who had a history of implementing the ASM model well. Thus, we did not study the average ASM apprenticeship, but rather what were identified as the better apprenticeships based on ASM nominations and prior evaluation data.

ASM had led us to believe that few alternative after-school activities were available in these schools and communities. However, we discovered that the overwhelming majority of control youth (91%) were involved either in an organized after-school activity (primarily) or paid work. *This changed our understanding of the experimental contrast in an important way:* instead of comparing ASM to no treatment, we were actually comparing ASM to the range of organized after-school activities in which these youth normally participate—in effect, an alternative treatment comparison.

Apprenticeships were located across 10 Chicago public high schools. Students selected an apprenticeship in which they were interested, and those applicants were randomly assigned to the apprenticeship or to the control condition (“business as usual”) by the research team. The majority of participants were African American (77%) and low income (92% received free or reduced price lunch).

We collected data on a diverse set of outcome variables representing four broad constructs: positive youth development, marketable job skills, academic performance, and problem behavior. Variables were assessed via multiple sources and methods: surveys of youth, interview ratings by human resource professionals in mock job interviews, archival school records, and observations by the research team. Both intervention and control youth were

assessed prior to the onset of the apprenticeship (pretest) and after the end of the spring apprenticeship (posttest). We analyzed the data using a hierarchical linear model (HLM) approach, controlling for key demographic variables and a pretest on the same variable as a covariate (when available).

Implementation

There were no major implementation problems aside from a high attrition rate, which is not uncommon among after-school programs. All programs followed the ASM model in having a strong emphasis on skill development. We considered a number of the apprenticeships to be exemplary, though three did not meet our expectations in terms of program quality.

As we learned from the Year One data the extent of extra-curricular participation by control youth, we changed our quantitative implementation measures so that youth assigned to the control group could rate the after-school activity in which they spent the most time. ASM youth rated their apprenticeship. Overall, ASM youth rated their apprenticeships as having significantly more key program design features that were identified as important to ASM programs, such as having an adult teach new skills and how to improve those skills, working on activities that are related to future careers and that are used or viewed by others, and having choice in the activity. However, apprenticeship and control group youth did not differ in their ratings of the social climate of their respective setting, suggesting that the control group after-school experiences constituted a strong alternative treatment.

Key impact findings from intent-to-treat analyses

In our primary analyses we estimated so-called “intent-to-treat” program impacts, which compared those who were assigned to treatment (regardless of whether they participated in the intervention) with those who were assigned to the control group (even though a few of the control youth managed to participate in an alternative ASM program). Intent-to-treat analyses preserve the integrity of the experimental design and are considered the most methodologically rigorous. They are also important for policy purposes as they address impacts regardless of implementation issues that affect participation rates.

We present first the results in terms of statistically significant results on specific variables. In brief, *there were statistically significant differences favoring ASM for positive youth development and problem behaviors, but no statistically significant differences between the groups for marketable job skills or academic outcomes. There were no statistically significant differences that favored the control group over ASM.*

- **Positive youth development.** Youth in the treatment group reported significantly higher self-regulation than youth in the control group. This reflected a preventive impact: both groups reported a decline in self-regulation over the course of the year, but the decline was less among ASM youth. There were no significant differences for the four other measures in this domain.
- **Marketable job skills.** There were no significant differences in this domain.
- **Academic outcomes.** There were no significant differences between ASM and control youth for school attendance or grades. There was a marginally significant difference favoring ASM on a scale that measured identification with school.
- **Problem behavior.** Youth in the treatment group reported significantly fewer problem behaviors than youth in the control group on a 10-item index. This again reflected a preventive impact: both groups reported a slight increase over time, but the increase was less for ASM youth. In terms of individual items, there were significant differences favoring ASM for two of the crime items: selling drugs and participating in gang activity.

Treatment-on-the-treated findings

We also wanted to determine whether ASM had an impact when we considered only those who actually received the treatment. Nearly half (47%) of the students assigned to the treatment condition dropped out. We therefore also conducted “treatment-on-the-treated” analyses. Treatment here was defined using ASM’s own dosage threshold: apprentices cannot miss more than 4 sessions per pay period (15 sessions) in order to receive a stipend (roughly 73% attendance). Therefore, youth who attended at least 73% of the sessions were defined as “treated.” Control youth who attended an alternative ASM program were excluded from these analyses.

As in the intent-to-treat analyses, there was a significant effect favoring ASM on self-regulation, but no effect on the other positive youth development variables. For marketable job skills, there was a significant difference (that was not found for intent-to-treat) favoring ASM on a composite index measuring student responses to mock job interview questions; however, ASM youth were not significantly more likely to be hired, nor were there other significant differences in this domain. In the academic realm, the scale measuring school identification was significant (it was marginally significant in intent-to-treat) and favored ASM, but all of the other outcomes remained non-significant. There was no significant difference on the problem behavior scale (which was significant in intent-to-treat).

Thus, the ASM effect on self-regulation was the same in both sets of analyses, whereas the problem behavior scale was significant only for intent-to-treat, and the school identification and one mock job scale were significant only for treatment-on-the-treated. On balance, given the policy importance of problem behaviors, we consider the intent-to-treat findings to be more positive for ASM and they are also based on a stronger research methodology. Thus, the strongest findings favoring ASM came from the strongest methodology.

Results for average effects per outcome domain

It is not unusual for treatment effects to be averaged within a particular outcome domain in meta-analyses or in reports such as those conducted by the What Works Clearinghouse. Such averages reflect treatment effects on broader constructs than any single measure. As ASM posits effects in broad outcome domains, this approach is appropriate for this evaluation. Given that there were a number of non-significant effects on specific individual variables, it is not surprising that the average treatment effects within domains are generally small. The weakest outcome findings, averaged by domain, were in ASM's two highest priority areas: positive youth development and marketable job skills.

Discussion

There are two contrasting interpretations of these data, each of which makes reasonable points. A positive perspective emphasizes that ASM was able to obtain significant positive results on important outcome variables despite several factors that worked against doing so (e.g., an alternative treatment control group, lack of substantial extra support for implementation) and

that these impacts can be meaningfully related to ASM vs. control experiences in their respective activities. A skeptical view of the findings emphasizes that few significant effects were found, effect sizes were generally small, and that testing a more representative sample of ASM instructors may well eliminate the few positive impacts that were found.

We believe that each of these perspectives makes reasonable points and that it is unwise to view this as a situation in which one needs to choose only one or the other. Few randomized evaluation studies have been done on after-school programs, especially for high school youth. High quality evaluation studies can provide information that is crucial to program development, for strengthening the ability of programs to produce strong effects. In turn, redeveloped programs need to be subjected to further evaluation, which can result in a cumulative process that greatly enhances program effectiveness. The after-school world is just at the beginning of this process.

Within this broader, historical context, we consider the ASM impacts to be promising. Although it is frequently the case that no significant treatment effects are found in experimental outcome studies, in this research ASM did have a significant impact in areas that are important to adolescent development and to policy. Moreover, it demonstrated these impacts in relation to what was essentially an alternative treatment comparison group. Nonetheless, we consider the counter-arguments from the skeptical perspective to be serious and these force us to view the outcomes with caution. The caution is with respect primarily to whether the average ASM apprenticeship (which we did not study) is likely to provide outcomes superior to what high school youth can obtain in alternative extra-curricular activities, after-school programs, and part-time jobs. The skeptical perspective may ultimately prove prescient with respect to what such an outcome study would reveal. In order to maximize the likelihood that the modal ASM apprenticeship proves superior in such a future outcome evaluation, it is important to improve the program model.

Pointing toward the future, we suggest that researchers study more carefully the experiences of control group youth and develop new measures of positive youth development. For programs, we suggest how ASM can learn from its own best practices in terms of strengthening its ability to engage youth and have youth produce high quality work. ASM also

needs to focus more on transferable knowledge and skills, both in terms of soft skills for the workplace and for knowledge, attitudes, and skills in the domain of positive youth development. To best enable youth to achieve key ASM outcome objectives, core program design features provide a promising foundation, yet important elements of what happens in the apprenticeships need to improve.

CHAPTER 1

OVERVIEW OF THE EVALUATION

This report presents results from a three-year, random assignment evaluation of After School Matters (ASM). After School Matters offers paid apprenticeship-type experiences to help adolescents begin to appreciate and adapt to the culture of the workplace and improve the “soft skills” increasingly demanded by employers. The major questions addressed by this evaluation are whether assignment to and participation in ASM apprenticeships results in gains in positive youth development, marketable job skills, and academic performance, and decreases in problem behavior.

This chapter begins with a review of the literature on prior after-school program evaluations. We also discuss the design of After School Matters programs and review prior literature on ASM. We conclude the chapter with key research questions for the evaluation.

In Chapter 2 we discuss the evaluation sample selection process, participant demographics, evaluation design, and analytic approach. In Chapter 3 we report results from our implementation measures comparing ASM programs to the control group extracurricular activities. In Chapters 4 and 5 we report impact findings from intent-to-treat/control comparisons and treatment/control group comparisons, respectively. The findings are discussed in Chapter 6.

Background

Over the past few years, after-school programs have attracted a strong and growing constituency among both academic theorists and policy makers. Challenging activities and relationships with caring staff can lead to important developmental gains by increasing skills, instilling confidence, broadening cultural horizons, and promoting positive values (Benson, 1997; Carnegie Corporation of New York, 1992; Connell, Gambone, & Smith, 2000; Hawkins, Catalano, & Associates, 1992; Hirsch, 2005; Lerner, 2004; Mahoney, Larson, & Eccles, 2005; National Research Council, 2002; Noam, Biancaosa, & Dechausay, 2003; Pittman, Irby, & Ferber, 2000; Quinn, 1999). Potential policy implications range from reducing youth crime

during the high-risk 3-6 PM period to supporting school reform efforts and promoting workforce preparation and positive youth development.

On balance, initial evaluation findings on the effectiveness of after-school programs are promising. Durlak, Weissberg, and Pachan (2010) recently completed a comprehensive meta-analysis on the effectiveness of after-school programs. Overall, they found an average positive effect size of approximately one-quarter of a standard deviation. The average effect was larger for programs they defined as being highly structured, programs whose activities were sequenced, active, focused, and explicit (S.A.F.E.).

However, there are important methodological limitations to much research on after-school programs. Many studies failed to include pretest measures of outcomes and had no or poorly matched comparison groups (Gottfredson et al., 2010). There have been few randomized controlled trials. Randomization is the best procedure for guarding against selection effects in which youth more likely to improve over time are disproportionately located in the treatment group. In that event, it is impossible to sort out whether effects are due to selection (who got into the program) or the program itself (the experiences of youth while in the program).

Furthermore, the large majority of programs evaluated using randomized control group designs served elementary and middle school students, rather than high school (HS) students (Gottfredson et al., 2010). The comparative paucity of studies on effects for HS students is not surprising as most after-school programs serve younger children. Although the 40 largest national youth organizations reach approximately 40 million youth, participation is extremely sparse at the HS level. In the Boys & Girls Clubs of America, for instance, only 10% of participants are ages 16-18. Indeed, throughout the western world, attendance in youth programs drops dramatically over the course of adolescence (Cotterell, 1996). Clearly there is a need to develop more attractive after-school programs for HS youth, which makes After School Matters, one of the country's largest programs for this group, an important program to study.

To the best of our knowledge, there is only one program that was unquestionably an after-school program, focused exclusively on HS students, and was evaluated via randomized experimental design: the Quantum Opportunities Program. In Quantum, youth in the

intervention group were expected to receive academic activities (e.g., tutoring), development activities (e.g., problem prevention, college planning), and community service. Case management, mentoring, and a stipend were also provided. Findings from the initial evaluation and a larger, follow-up evaluation indicated a significant or marginally significant positive effect on HS graduation and greater likelihood of enrollment in some type of post-secondary schooling, but no impact on test scores or problem behaviors (Hahn, Leavitt, & Aaron, 1994; Maxfield, Schirm, & Rodriguez-Planas, 2003). Both evaluations reported major implementation problems (see also Maxfield, Castner, Maralani, & Vencill, 2003), as most youth did not receive the specified dosage and several components were not implemented as fully as specified. In addition, almost all youth served were female.

The other random assignment evaluation of a HS after-school program involved the Children's Aid Society-Carrera Program (Philiber, Kaye, Herrling & West, 2002). Targeting students in both middle and high school, the program included several components: vocational, academic, family life and sexuality education; arts; sports; mental health counseling; and medical care. Extensive implementation data were not reported. Regarding outcomes, findings indicate that intervention females had significantly fewer pregnancies than control youth. There was no significant impact on males' sexual behavior. There were no differences for either gender in grades, delinquency, or drug outcomes. Thus, this program has some promise in terms of preventing pregnancies among girls, but it is of unclear value otherwise.

Thus, among the few HS after-school programs that have been developed and experimentally evaluated, there are some promising findings, but the range of outcomes appears limited, especially for males. Implementation data were available primarily for one program, and in that instance (Quantum) the problems were extensive. Neither program had a strong orientation toward providing work or apprenticeship experiences.

In addition to formalized programs such as Quantum Opportunities and the Children's Aid Society, the after-school landscape includes a wide range of extracurricular activities: school- and community-based sports teams, art and music ensembles, academic and hobby clubs, faith-based youth groups, and opportunities for community service. A majority of American high school students (approximately 70%) are involved in at least one such activity, and many

participate in two or more (Feldman & Matjasko, 2005; Mahoney, Vandell, Simpkins, & Zarrett, 2009). Although there are potential drawbacks to structured activities (e.g., the relationship between sports and alcohol use; Eccles & Barber, 1999), participation is generally associated with positive outcomes across multiple domains. These include higher educational performance and attainment, reduced delinquent and high-risk behavior, and greater psychological wellbeing such as increased self-esteem (Bartko & Eccles, 2003; Eccles & Barber, 1999; Feldman & Matjasko, 2005; Mahoney et al., 2009; Pedersen & Seidman, 2004). However, findings from studies of extracurricular activities must be interpreted in light of their methodological limitations. Most notably, we are not aware of randomized controlled trials in this area, indicating the persistent threat of selection effects.

Because After School Matters features apprenticeship experiences, at this point we will turn our literature review to consider findings regarding the value of work experiences in adolescence and, then, current programs for HS students that are oriented toward the world of work.

In his seminal work, Erikson (1968) argued that adolescence is a period during which important attitudes toward work and vocational identities are developed. One important marker of adulthood is taking up a steady full-time job and adolescence can be viewed as a preparatory period in which youth gain information about jobs in anticipation for the future (Mortimer, 2003). Part-time work can generate positive outcomes for youth, including increased confidence and time management ability, enhanced academic success, and later life advantages that spring from expanded networks (Mortimer, 2003; Mortimer, Pimentel, Rye, Nash, & Lee, 1996). Teenage work can increase wages, employment, and occupational status up to 10 years later (Carr, Wright, & Brody, 1996; Ruhm, 1997). Among minority youth, those who do not work at all during HS are at the highest risk of dropping out (Tienda & Ahituv, 1996). For youth from low-income families, jobs can offer an alternative to patterns of neighborhood crime and unemployment and enable them to obtain skills and resources that may apply to better jobs in the future (Newman, 1999). Employment experiences can help youth avoid the dead-end jobs that characterize the secondary labor market; those who miss good jobs early may be permanently tracked onto inferior job trajectories (Ellwood, 1982). Jobs that are more challenging and

develop skills are the most valuable (Mortimer, 2003), though many jobs for young people do not provide these types of experiences (Greenberger & Steinberg, 1986). Moreover, some data suggest that particularly intensive work styles (>20 hours/week) are associated with short-term delinquent behaviors, though this does not appear to increase adult criminality and substance use (Staff, Messersmith, & Schulenberg, 2009).

Given the value attached to working, there have been a number of efforts over the years to improve workforce development of young people. For example, large federal sums were committed to such programs during the Carter presidency (Youth Employment and Demonstration Projects Act). The hasty start-up, unfortunately, did not result in many high quality evaluations, particularly in areas closest to our concerns, such as programs focusing on labor market preparation for in-school youth (National Research Council, 1985). The School-to-Work Opportunities Act of 1993 was another such effort.

The most intensive, current effort to improve workforce development of in-school youth focuses on the creation of career academies within high schools. These schools are organized around a career theme, and provide academic and technical curricula, work-based learning opportunities, as well as links to local employers. Findings from MDRC's evaluation of 9 academies, using an experimental design, suggest a number of positive outcomes (Kemple & Willner, 2008). In particular, male academy students earn more than their non-academy counterparts, with the highest impact among those most at risk for dropping out. Such gains were not evident among female students. Additional problems were noted, including a high attrition rate and a reduction in academic course taking. This type of whole-school reform is also costly and complex, making it difficult to implement well (National Research Council, 2003).

This literature review suggests that after-school programs, part-time work experiences, and workforce preparation programs can have positive effects on HS students. After-school programs that have an apprenticeship orientation have the potential to provide the benefits of successful part-time work experience at a lower cost than many workforce preparation programs. Moreover, as an after-school program, they have the latitude to focus more broadly on positive youth development than might be the case with programs targeted exclusively at workforce preparation. After School Matters is just such a program.

After School Matters

After School Matters provides apprenticeship-type experiences to students in Chicago public high schools. ASM began in 2000 as an outgrowth of the Gallery 37 program, a City of Chicago initiative focused on apprenticeships in the arts. Last year, ASM was located in 65 Chicago public high schools. In the fall of 2009, there were 305 apprenticeships across the city, which enrolled approximately 7,400 HS students. We believe ASM to be the largest, one-city after-school program for HS students in the country. In this section, we will first describe the ASM program, discuss its rationale in terms of developmental theory and intervention theory and then present findings from prior research on ASM.

An apprenticeship lasts for 10 weeks in the fall and 10 weeks in the spring (all students in the evaluation participated in the same apprenticeship in both the fall and spring). The apprenticeships meet for 9 hours/week: 3 hours after school, 3 times weekly (180 hours for the year). During and prior to our research, students were paid \$900 (equivalent to \$5/hour) for participating in both sessions, with adjustments for attendance (subsequent to our study, the stipend was lowered due to budgetary considerations).

Each apprenticeship is organized around a particular enterprise, which range widely. Examples include those focused exclusively on technology, such as web design or computer repair; those which combine technology and art, such as producing social documentaries; those which are primarily artistic, such as improvisational theater groups or painting; and those which have a sports orientation, such as lifeguarding or learning how to teach young children to play soccer. Each apprenticeship session involves work in the designated area, learning and making use of relevant skills to accomplish a task. Instructors are present to provide information, guidance, and feedback, and introduce students to the standards, language, and culture of that line of work. The apprenticeship often culminates in a final product or performance. Along the way, there are often intermediate products designed to provide success experiences that serve to solidify motivation and mark increasing levels of skill development that will be utilized in the final project.

Two paid instructors direct each apprenticeship. The instructors (who are generally not teachers) have expertise in—and in many instances earn their livelihood through—the activity that is the focus of the apprenticeship. ASM provides both beginning and advanced training sessions for instructors. There are regional directors overseeing programming in each of 6 geographical divisions. Within each region, there are ASM specialists assigned to specific schools to provide additional training, supervision, and logistic assistance. Each school had a paid school staff member who serves as a liaison between the school and its ASM programs. The great majority of school-year apprenticeships take place at the host high school.

ASM experiences bear a familial-like resemblance to other apprenticeship-type interventions that provide exposure to work skills or environments in a particular occupation (see Halpern, 2009, for a recent study of apprenticeships in the American context). They do not provide the kind of intensive, on-the-job training in technical trade skills such as are found in Germany (see Hamilton, 1990). ASM youth obtain training in technical skills within a positive youth development context that enables them to begin to appreciate and adapt to the culture of the workplace and develop the “soft skills” (e.g., Murnane & Levy, 1996) increasingly demanded in the 21st century economy.

We now turn to consider the developmental and intervention theory rationale for possible ASM effects.

Developmental Theory Rationale

Erikson (1968) characterized adolescence as a developmental period with an overarching concern with identity development. Adolescents develop a heightened concern with who they are and who they will be in the future. Identity development has a strong contextual component, as young people consider how they will fit into and find a valued place in the adult world. Erikson’s approach is quite consonant with more recent cultural perspectives on development. Rogoff (2003), in particular, conceptualizes development in terms of transformation of participation in cultural activities. Creating an identity—a place for oneself—in the world of work is a fundamental task of identity construction in our culture. This involves a variety of subtasks, such as exploring different possible kinds of work to see what one enjoys and is good

at, clarification of occupational values, increasing one's sense of self-efficacy, and developing skills in problem solving, self-regulation, and communication with adult authority figures.

ASM apprenticeships may provide these varied developmental experiences. They allow HS students the opportunity not only to explore a specific line of work, but probably even more importantly, to also acquire a taste of the culture of the world of work and gain an appreciation of core, generic features of work life. Apprenticeships may provide the opportunity to learn the value of hard work, to work on challenging tasks, have high standards, work effectively as a member of a team, develop constructive relationships with a supervisor/boss, and so on. These are critical experiences that are at the heart of adolescent development and provide a basis for imagining—and taking steps to achieve—a positive future.

The opportunity to engage in adult-like activities, to be exposed to the world of work (including working professionals), and to be paid for doing so, just as they would in a job, likely provides much of the motivation for student enrollment. These factors can also be used by apprentices as legitimate arguments should they feel the need to justify their enrollment to any dubious peers.

Intervention Theory Rationale

We draw on three literatures which suggest that ASM may well incorporate principles of strong intervention design: community psychology, youth development as represented by the Eccles et al. National Research Council (2002) report, and design theory from education.

From community psychology, Seymour Sarason's work (especially Sarason, 1972, 1982) focuses attention on how the culture of a setting influences the effectiveness of change efforts. The culture of the apprenticeship, and the social regularities that characterize it (Seidman, 1988), should be aligned with program goals. From what we know of ASM from published reports (Halpern, 2006; Larson, 2007), good apprenticeships appear to provide a culture where hard work is expected and valued, where cooperative teamwork is cultivated, and where skill development and task completion are rewarded.

As part of its report on youth programs, the NRC panel (2002) identified features of positive developmental settings that have justifiably attracted considerable attention. The panel

reviewed a substantial body of research across disciplines to develop this list. Based on prior research, good ASM apprenticeships seem to rate highly on most of those dimensions. Of particular importance in terms of achieving program goals, good apprenticeships are likely to rate highly on appropriate structure, supportive relationships, opportunities to belong, positive social norms, support for efficacy and mattering, and opportunities for skill building. If strong settings are created along these dimensions, the apprenticeships could be a potent intervention.

Finally, ASM specifies that apprenticeships should adhere closely to a number of important design principles: youth work on authentic tasks that have meaning in the “real world” and they learn by doing (Edelson, 2001; Schank, 1995); the tasks are challenging, in that they are often slightly beyond youth’s present ability yet manageable with assistance, what Vygotsky (1978) would refer to as the zone of proximal development; youth are encouraged to take initiative, make decisions, or be creative (Barron, 1998); youth are provided opportunities to teach or share what they have learned with their peers (Brown & Campione, 1996); instructors provide scaffolding and guided feedback (Jackson, Krajcik, & Soloway, 1998; Rogoff, 1990); tasks and skill are developed sequentially so that proximal goals map logically onto distal objectives and the final product/performance demands consideration of everything (or almost everything) that they have learned (Brown & Campione, 1996). Many of these design features are rated by ASM supervisors during site visits, and have been found to characterize good apprenticeships by prior researchers (Halpern, 2006; Larson, 2007).

Prior Research on ASM

Existing research suggested that After School Matters was ready for an experimental evaluation. ASM exit surveys, completed anonymously by youth over the Internet, revealed widespread youth enthusiasm for the programs. Over 80% of youth respondents indicated that ASM improved their ability to set a goal and work to achieve it; get things done on time; communicate clearly; and work with others on a team or group project. When asked to choose a word from a list of 25 possibilities that best described how they felt while in the program, the most frequently chosen descriptors were excited, comfortable, interested, and challenged, while the least frequently chosen were disrespected, unwelcome, lazy, and angry. ASM instructors also rate youth on varied skill domains, and although the data have not been analyzed statistically

(and there is no control group), the ratings suggest that the instructors perceive skill gains in a variety of domains.

A number of ASM programs have been studied by qualitative researchers. Larson (2007) studied a media apprenticeship that focused on skills in using computer software and video equipment. His research revealed that students developed strong skills in teamwork as part of that experience. More generally, good ASM apprenticeships appear to correspond in many respects to the broader range of high quality after-school programs that the Larson research group is studying (see Larson et al., 2004).

Observing 24 apprenticeships over the course of 2 years, Halpern (2006) described the strong design features of ASM, identified earlier in this chapter. He found that many youth appeared to make notable gains from program participation in the areas of discipline-specific knowledge and skill, general executive skills, social and interpersonal skills, and self-development. At the same time, he noted that there was variability in instructor quality and in program attendance (dosage). Furthermore, he noted that the typical instructor estimated that apprenticeships result in significant gains for only about 20-25% of youth (Halpern, with Kimondo, 2005). It is not clear from this estimate what change would be seen averaged across all participants.

In a quasi-experimental study, Goerge, Cusick, Wasserman, & Gladden (2007) found ASM was associated with improved school attendance and fewer failing grades of F, and that students who have participated in four or more semesters of apprenticeship were 2.4 times as likely to graduate HS. There are a number of apparent methodological limitations to this study that suggest that the results be viewed with some caution; furthermore, given insufficient detail, we are unclear about other important aspects of the methodology. There is no indication in the report that any matching procedure (e.g., propensity scores) was utilized to help select an equivalent control group. Instead, the control group appeared to be constituted by all students in the school who did not participate in ASM. If so, the control in effect pooled together students who were involved in an alternative extra-curricular activity (a reasonable comparison group) with those not engaged in any extra-curricular activity (a less suitable control group), rather than separate analyses to compare each group to ASM. It is not surprising in this context that there

were differences in students who signed up to participate in ASM compared to students who did not. Prior to signing up, the students who were interested in ASM were already more likely to have higher attendance than the control group. Of those students who initially expressed interest in ASM, more than half of the students did not go on to participate; and those who did participate were students with the fewest failing grades and fewest school absences in the prior year. There may also have been differences between the groups on important variables that were not measured (and therefore could not be controlled for statistically). Without an appropriate comparison group, it is difficult to interpret the evaluation results, as positive effects found in the evaluation may be due to participant self-selection. These limitations suggest the need for a random assignment study.

Overall, evaluations of ASM have shown potential, yet they are inconclusive. The findings suggest that some youth benefit, but whether enough do so sufficiently to create an average treatment effect is unclear. Therefore, given the importance of ASM, we conducted a rigorous evaluation employing random assignment and assessed multiple outcome domains.

Research Questions

This evaluation investigates whether assignment to and participation in ASM results in gains in positive youth development, marketable job skills, and academic performance, as well as reductions in problem behaviors. Conceptually, the development of marketable job skills can be considered as part of positive youth development for HS students; our separation of these is primarily for heuristic purposes. As an organization, ASM is most concerned with outcomes in the areas of positive youth development and marketable job skills. ASM does not emphasize academic or problem behavior outcomes, but given their importance to the field, neither ASM nor the investigators wanted to ignore these.

Gains in Positive Youth Development

In this domain we include important values, attitudes, and skills that should change as a result of apprenticeship experiences that involve making integral contributions to successful work products/performances. These changes should be of general developmental benefit, as well as providing human and social capital that can be of specific use in the workplace. Our focus is

on self-efficacy, self-regulation, occupational values, and relationships with adult authority figures and with peers.

Self-efficacy is the belief in one's own capacity to marshal the motivation, cognitive resources, and actions required to navigate challenging situations (Wood & Bandura, 1989). The belief in one's own ability to exercise a degree of control over life circumstances is critical because it influences "how people think, feel, motivate themselves, and act" (Bandura, 1995, p.2). Self-efficacy influences the goals students set for themselves (for example, students with a higher sense of self-efficacy establish increasingly challenging and difficult goals) and the commitment with which they pursue those goals (Zimmerman, 1995b; Zimmerman, 1990). Self-efficacy beliefs are impacted by four main forms of influence: (1) Mastery experiences, which provide direct evidence of having what it takes to succeed at a task or challenge; (2) Vicarious experiences, which provide evidence based on observation of others that the young person can succeed if others with similar characteristics can do so; (3) Social persuasion experiences, which provide opportunities for youth to be convinced verbally that they are capable of success; and (4) Positive physiological/emotional experiences, which generate higher self-assessments of capability than negative experiences (Bandura, 1995). Educational experiences that provide these experiences should lead to self-efficacy gains (Bandura, 1997), and ASM apprenticeship experiences may do so as well.

Self-regulation describes how individuals pursue goal-directed activities across changing situations and over time through the management of their attention and emotions (Karoly, 1993; Zimmerman, 1995a). Persons with high self-regulation are aware of their emotions and able to control the duration and intensity of the attentive distractions they produce, while individuals with low self-regulation may have difficulty disengaging sufficiently from their current emotional state to direct their attention to the task at hand (Luszczynska, Diehl, Gutierrez-Dona, Kuusinen, & Schwarzer, 2004). Self-regulation is an important part of how students handle their own learning processes when they experience distractions from their work (Corno, 1993; Karoly, 1993). Additionally, self-regulation is an important part of adolescents' involvement in work and career activities (Owens & Schneider, 2005). Self-regulation is quite salient to youth from these types of neighborhoods, which have high rates of poverty and crime and can be quite stressful. It

will be evaluated whether ASM provides a context within which youth gain increased skill in self-regulation.

Occupational values are the beliefs and attitudes that individuals develop toward work that influence vocational striving and work choices in life (Mortimer, 2003). Occupational values have both intrinsic dimensions (e.g., autonomy, compatibility with one's interests and abilities) and extrinsic dimensions (e.g., money, social status). The quality of a work experience can have an impact on the formation of values about work. For example, findings from a multi-wave longitudinal study of youth and occupational values (Mortimer et al., 1996) showed that the chance to learn useful skills on the job had a strong positive influence on the development of occupational values, which, in turn, predicted subsequent job experience. Students with stronger values were more likely to report better job experiences, suggesting both that their own orientation to work may help youth find better jobs and that they engage with the job in a manner more likely to produce a satisfactory experience. Mortimer (2003) also found that young adults whose high school jobs provided opportunities for learning new skills were much more likely to be employed seven years later in jobs they considered career related. Mortimer et al. (1996) concluded that the quality of adolescents' work experience—especially the opportunity to learn new skills and manage challenges—is the critical factor in how youth employment will ultimately impact occupational values. ASM apprenticeship experiences can involve learning challenging new skills and therefore may lead to increases in occupational values.

As young people enter the workplace, one of the most critical challenges they face is to develop a satisfactory relationship with their supervisor (the “boss”). Drawing on the concept of internal working models of relationships from attachment theory (e.g., Bretherton & Munholland, 1999), it is possible that the relationship that youth develop with their ASM instructors can provide them with a positive working model of the kind of relationship that they can expect with their eventual workplace supervisor. Although relationships with ASM instructors would not have the kind of affective, familial-like quality that occurs among early adolescents at good after-school programs (Hirsch, 2005; Hirsch, Deutsch, & DuBois, 2011), prior research has shown they do mimic and anticipate quite effectively the kinds of relationships that youth will encounter in the workplace (Halpern, 2006). Effective relationships with ASM

instructors should provide apprentices with a more positive orientation toward the kind of relationship they can eventually establish at work, approaching the potential supervisor as a potential mentor (for the value of mentoring relationships to youth, see DuBois & Karcher, 2005; Hamilton & Hamilton, 2004; Rhodes, 2002).

In addition to forming relationships with adults, teens in ASM programs are required to collaborate with other students in accomplishing apprenticeship related goals. Experience working with other youth provides students with opportunities to learn skills related to social development and is a critical component of adolescent developmental processes (Hansen, Larson, & Dworkin, 2003). Whether ASM provided more opportunities to develop interpersonal social skills will be investigated.

Gains in Marketable Job Skills

ASM apprenticeships are designed to provide enrichment experiences and career exploration in terms of the specific foci of the apprenticeship (e.g., web design, theater production). These are valuable experiences and an important motivation for participation for many young people. However, given the modest labor market in many of these domains, and the limited technical skills acquired through ASM experiences (Halpern, 2006), if we restricted our conceptualization of marketable job skills to technical expertise in a specific labor market, the effect of ASM is likely to be limited as well. But that is neither necessary nor appropriate. Current conceptualizations of marketable job skills, particularly for entry-level positions, have broadened considerably in ways that map well onto potential ASM outcomes.

Beyond academic preparation and occupation-specific skills, a variety of generic work dispositions and skills have received increased recognition as vital for the 21st century economy. Often referred to as “soft skills” (Murnane & Levy, 1996), evidence for their importance has come from surveys of employers (e.g., Holzer, 1996; Rosenbaum, 2001), reviews of empirical research on the employment interview (e.g., Eder & Harris, 1999; Huffcutt, Conway, Roth, & Stone, 2001), and advocacy briefs from the business community (e.g., Corporate Voices for Working Families, 2006). We found similar views among a group of senior human resource professionals we consulted about the local, Chicago labor market. The requisite characteristics

include communication skills, the ability to work collaboratively in teams, problem solving, goal orientation, self-confidence, accountability and responsibility. These facilitate the opportunity to participate effectively in the culture of the contemporary workplace. Many employers assert that these are the skills they look for in entry-level employees; they fully expect to train them in technical skills, which are often quite specific to their own workplaces.

ASM apprenticeships offer opportunities to learn these soft skills. The skill, teamwork, and final product emphases should result in positive ASM effects for these types of marketable job skills and, in turn, lead to positive hiring recommendations for entry-level positions.

Gains in School Attachment and Academic Performance

In urban high schools approximately only half of students who enter go on to receive a high school degree (National Research Council, 2003). This is true in Chicago where, according to findings from the Consortium on Chicago School Research (Allensworth, 2005), only 54% of entering freshmen graduate four years later, and where some schools served by ASM have graduation rates of only about 30%. In this context, the most important educational outcome would appear to be student retention and, ultimately, HS graduation. Progress toward this goal would be marked by improved attendance (Rumberger & Larson, 1998) and grades. The National Research Council (2003) identified a number of factors that are likely to have positive effects on urban HS students: challenging instruction and support for high standards; choices that make the curriculum relevant to adolescent long-term goals; and promoting a sense of belonging by creating a supportive social context. These factors map well onto the experiences that ASM apprenticeships purport to offer youth in their high schools. Specifically students who participate should report greater attachment to school through both a sense of belonging as well as a greater perception of the instrumental value of education. As a hard measure of academic performance, we will test whether apprentices demonstrate increased attendance and better grades.

Decreased Problem Behavior

The Eccles et al. NRC report (2002) concluded that community youth programs decreased rates of problem behavior. There was some question as to whether after-school programs in particular had such an effect (Gottfredson, Gerstenblith, Soule, Womer, & Lu, 2004). However, the recent meta-analysis by Durlak & Weissberg (2010) did find a positive effect for after-school programs on problem behavior for the more focused programs. Moreover, as previously discussed, the NRC report identifies features of positive youth settings that likely account for positive outcomes and ASM would appear to load highly on most of those features. Thus participation in ASM should result in reductions in rates of problem behavior.

CHAPTER 2

RESEARCH METHODS

The research involved a randomized controlled trial using a mixed methods assessment approach. Youth who applied for a specific apprenticeship were randomly assigned to either the intervention (apprenticeship) or control group. Both intervention and control youth were assessed prior to the onset of the apprenticeship (pretest) and after the end of the spring semester, when the apprenticeship ended (posttest). Background, implementation, and outcome measures were assessed via multiple sources and methods (see **Table 2.1** for an overview). We studied a total of 13 apprenticeships and their respective controls yielding a total of 535 participating youth. HLM analyses compared treatment and control group means for each outcome measure.

Sample Selection

Apprenticeships

Selection of apprenticeships focused on those instructors who had a history of implementing the ASM model well. We were concerned with program fidelity in terms of instructors having a skills orientation and sought especially to avoid instructors (particularly in the arts) who might be charismatic but not adhere strongly to the program model. ASM program directors were asked to nominate a number of the best programs from their region according to that criterion. Final programs selected for the evaluation had site visit ratings (conducted by ASM specialists), data from ASM youth exit surveys (completed by former apprentices), and program attendance records that were reasonably consistent with our criterion. All instructors had at least 1 year of ASM experience. Different apprenticeships were chosen each year. There was reasonable variation in the substantive foci of apprenticeships.

Some instructors who were nominated declined to participate and some other instructors who agreed to participate withdrew prior to program onset due to personal or family reasons. The final group of apprenticeships was considered among the better ones in ASM but not the very best exclusively.

ASM instructors involved in the research were not provided with feedback or supervision beyond what they would ordinarily receive. Apprenticeships in the study received priority in obtaining necessary supplies in a timely manner. ASM also provided an additional workshop for apprenticeships involved in the evaluation. The workshop taught apprentices how to present themselves effectively through résumés and job interviews.

Schools

Each year, the apprenticeships in the evaluation were located in 3-4 Chicago public high schools. Three apprenticeships had taken place at the same school in the year prior to participating in the evaluation. The rest of the apprenticeships were placed at different schools during the evaluation year for pragmatic purposes. We selected schools with a positive history of involvement with ASM, which included sufficient youth enrollment and a supportive principal.

Among the 10 participating schools, the four-year average graduation rate was 57% (ranging from 30 to 69%). The average percentage of students passing the 11th grade statewide achievement tests was 18%, with large between-school variation (passing rates ranging from 3% to 40%). Most schools in the study had a large number of low-income students, with an average of 86% of the study body receiving free or reduced priced lunch. All schools had majority non-white populations; seven schools had a majority African American student body (average 83% African American). Three schools had a majority Latino student body (average 58% Latino).

Youth

Recruitment. As per ASM's usual practice, an information fair staffed by apprenticeship instructors was held in each school in late September. At this time, youth signed up for a specific apprenticeship, completed our pre-apprenticeship computer survey, and were required to return signed parental consent/youth assent forms. Typically, youth sign up for one semester at a time. For the purposes of the evaluation, students were required to sign up for both Fall and Spring semesters at the time of recruitment. In initial talks with ASM, they believed that the majority of youth who sign up in the Fall continue through Spring. Additionally, they were interested in moving toward offering year-long apprenticeship programs. Both ASM and the

research team believed two semesters of the intervention would give a better chance of detecting program impacts than one semester. It should be noted that there is no limitation to participation in the program over the course of high school and it is always possible that an even greater dosage would be associated with improved outcomes.

After the initial sign up, instructors interviewed students to determine eligibility. Normally, instructors are able to select the final group of students for their program. Instead, for purposes of evaluation research, instructors were asked to select *out* any student who would not be acceptable for their program. Those students were ineligible for the study. However, instructors were not allowed final choice of apprentices. Allowing instructors to choose their apprentices per normal would violate our random assignment evaluation and present a serious challenge (possible selection effect) to the validity of the study. Youth who completed required enrollment procedures and were deemed eligible by instructors were randomly assigned to the intervention (apprenticeship) or control condition (“business as usual”) by the research team using a random number generator in SPSS.

In the first year, we selected up to 50 interested youth per apprenticeship (25 intervention and 25 control youth) who completed the qualification procedures. We increased our targeted program enrollment to 56 per apprenticeship (28 intervention and 28 control) in year 2 and to 60 per apprenticeship (30 intervention and 30 control) in year 3, in order to address treatment attrition (as described in the following section). Any additional youth who signed up but did not complete enrollment or who were rated as ineligible by the instructors were not entered into the random assignment lottery (i.e., they were not part of the study).

Although *all* subjects were randomly assigned, if an apprenticeship did not reach the target enrollment at the time of the lottery, we instructed the SPSS program to assign more subjects to the treatment group than the control group. In some instances, when there were a large number of no-shows after the initial lottery, we conducted a second lottery. For the second lottery, we took all of those initially assigned to the control group and had the SPSS program randomly select X number of youth (no more than 5), who were then put into the treatment group. These procedures maintained the integrity of the random assignment, experimental design, while also enabling the apprenticeship to meet ASM enrollment requirements. ASM

became very concerned with meeting enrollment targets organization-wide during the second and third years of the study.

Youth assigned to the control group were informed that they could not join any other ASM apprenticeship for the duration of the school year. Control youth who wished to participate in the following summer (after the posttest assessment) were given priority in enrollment procedures for summer programs. Throughout the evaluation, research staff monitored compliance with group assignment during ASM session observations and using ASM enrollment registers.

In the first year of the study, two of the three apprenticeships remained in the same location as the previous year, as is often the case with ASM. However, it became apparent that many students who participated in the apprenticeship in prior years attempted to reenroll. In effect, these students would have had two (or more) years of the assessed apprenticeship and their inclusion would compromise the evaluation design. In order to accommodate returning students who wished to participate, programs were allowed to enroll up to 5 “veteran apprentices,” but they were not considered research subjects. In the following years, apprenticeships were placed in different schools (with fewer alternative ASM apprenticeships) in order to avoid enrolling students with extensive prior ASM participation.

Attrition. On the basis of data presented to us by ASM and discussed with them, we had assumed that the attrition rate would be low (2-3 students/apprenticeship). However, this proved to be a large underestimation. In fact, nearly one half of students assigned to treatment dropped out before the end of the apprenticeship (see **Table 2.3**)¹. In order to minimize the attrition problem, we increased our targeted enrollment figure from 25 to 28 assigned to treatment per apprenticeship in the second year, and 30 per apprenticeship in the third year. We also developed stronger enrollment procedures than is typical of ASM. In addition to the normal posting of lists and public address announcements in the school, we had instructors phone all students admitted to the apprenticeship to inform them of their selection. Second phone calls were made in the spring to remind apprentices to return for the second semester.

¹ Although levels of attrition were higher than anticipated, these figures are similar to attendance levels reported in evaluations of other after-school programs (e.g., Gottfredson et al., 2010).

Participants

Final analyses were limited to students with outcome measures in at least one domain (i.e., participated in the posttest assessment or we were able to obtain a school transcript). This resulted in a total of 535 students in the study (304 assigned to treatment, 231 assigned to the control group).²

Participation in our posttest assessment (computer survey and mock interview) was 86.1% (see **Table 2.4**). Youth assigned to treatment (88.8%) were more likely to complete the posttest survey than those assigned to the control group (82.5%; $\chi^2 = 4.50$, $p = .03$).

Characteristics of the Entire Sample

Participants were Chicago public high school students. **Table 2.5** presents demographic characteristics for the entire sample. They were primarily 9th (36%) and 10th graders (30%) and were an average of 15.9 years old at the pretest survey. Similar to school demographic information, the majority was African American (77%) and low income (92% received free or reduced price lunch). The mean composite standardized test scores (expressed in stanines) for the sample was 4.88.

The majority of participants were involved in an extracurricular activity in the year before the study (82%). As stated earlier, students who had been in the same apprenticeship in the prior year were not allowed to be a part of the study. However, students who had been part of a different ASM program (either a club or another apprenticeship program) were allowed to participate and several had prior ASM experience (21%). Finally, 75% of students reported having a prior part-time job (e.g., babysitting, yard work, fast food restaurants, grocery/convenience stores, etc.).

As would be expected with random assignment, there were no significant pre-test differences between those assigned to treatment and those assigned to the control group. **Table**

² The assignment of more subjects to the intervention than the control group, as described previously, was the principal factor leading to the differential size of the groups. The differential assignment was modest. For example, consider that if we had changed the assignment ratio so that on average, per apprenticeship, we assigned 3 fewer youth to the treatment group than to the control group (relative to our actual assignments), the resulting size of the treatment group would have been 265 and the control group size would have been 270.

2.5 also lists background demographics for assigned-to-treatment and control group youth as well as F-test/Chi Square statistics for mean group differences.

Treatment Group Determination

Intent-to-Treat Group

First, we completed analyses examining the effect of treatment assignment using the entire participant sample. Treatment was defined as having been assigned to the apprenticeship. In addition to including all those who completed treatment, it included those who dropped out (received no or a low dose of the intervention), four of whom subsequently enrolled in an alternative ASM apprenticeship. In addition, we included in the control group, for these analyses, 24 youth who, despite our best efforts, managed to enroll in another ASM apprenticeship.

Although defining the treatment and control groups in this way may seem counterintuitive to some in the practice community, many evaluation researchers privilege findings from intent-to-treat analyses. This is because drop-outs from the treatment or control group threaten the randomization design. In particular, there is the worry that those who remain in treatment would be most likely to make gains under any circumstance (a form of selection effect). In addition, an important policy question is whether the treatment is effective for the entire group targeted to receive it, as public resources are spent on the entire group, not just those who remain in treatment or attend frequently. Moreover, if there is a significant treatment effect, policy makers can be assured that there will be positive gains regardless of implementation issues that may affect the participation rate.

Treatment-on-the-Treated

As revealed by high attrition rates, many apprentices did not receive a full dose of the treatment. ASM payment guidelines state that apprentices cannot miss more than 4 sessions per pay period (15 sessions) in order to receive a stipend (roughly 73% attendance). Using this guideline, students who attended at least 73% of the sessions were considered to have received the intervention and were included in the “treatment-on-the-treated” group. Participants assigned to treatment who did not meet this attendance standard were excluded from these analyses. Demographics for the sample for treatment-on-the-treated analyses are presented **Table 2.6**.

They were an average age of 15.9 years, 79% were African American, 59% were female, and 93% received free or reduced price lunch. There were no significant differences between the treatment and control groups on any background demographic variables. We note that these analyses excluded 4 treatment drop-outs who enrolled in an alternative ASM apprenticeship, as well as 24 control youth who enrolled in an alternative ASM apprenticeship.

Table 2.7 displays a comparison of those who met the ASM standard for treatment dosage with those who were assigned to treatment but dropped out. Youth in the treatment-on-the-treated group were more likely to be African American (83% v. 71% of dropouts; $p < .05$), whereas those who dropped out were more likely to be Latino/a (17% of treatment-on-the-treated v. 28% of dropouts; $p < .05$). Those in the treatment-on-the-treated group were also marginally more likely to receive free or reduced priced lunch compared with those who dropped out ($p < .10$). On measures collected at the pretest survey, youth in the treatment-on-the-treated group rated themselves as having significantly higher self-efficacy ($p < .05$) and marginally higher self-regulation ($p = .10$) compared to those who dropped out.

Control Group Extracurricular Activities

When we began this research, ASM told us that there were few alternative after-school activities available at these schools (and in these communities). Accordingly, we expected few members of the control group to be engaged in alternative activities and thus to constitute a no-treatment control. However, we discovered that the overwhelming majority of control youth (91% across all three years) were involved either in an after-school program (primarily) or paid work.³ In response, in Years 2 and 3 we asked the control group to rate their most time-intensive extracurricular activity using our implementation assessment measures. The distribution of control youth's most time intensive activity is found in **Table 2.8**.

³ In the final year (Year 3) of data collection, we undertook an additional assessment to check the accuracy of this figure. When students came to the part of our web assessment that assessed this domain, a trained research staff member conducted a brief, individual interview. The interviewer assessed whether the student had engaged in an organized activity that took place outside of regularly scheduled school classes, occurred during the specified time frame, and was led or supervised by an adult. The interview-based assessment led to only a modest decrease in the participation rate (75%) compared to the figure that was obtained from the survey questions (84%) for that cohort.

More importantly, this changes our understanding of the nature of the experimental contrast. Instead of comparing ASM to no treatment, we are actually comparing ASM to the range of organized after-school activities in which these youth normally participate—in effect, an alternative treatment comparison.

Data Collection/Research Procedures

Data Sources

Variables were assessed via multiple sources and methods: surveys of youth, interview ratings by human resource professionals in mock job interviews, archival school records, and observations by the research team.

Time of Assessments

All measures of positive youth development were obtained via a web-based computer survey that was given as part of the application process prior to selection into the intervention or control groups (pretest) and at the end of the spring semester, when the apprenticeship ended (posttest)⁴. Due to the logistics and expense involved, the Mock Job Interview was conducted only once, after the end of the spring apprenticeship. Grades and other school records were obtained for each semester of the evaluation year. Self-report problem behavior measures were obtained as a part of the computer survey at pretest and posttest.

We took several steps to make sure that youth completed assessments. The pretest assessment was a part of the required application procedure for the apprenticeship. For the posttest assessment in the spring, repeated PA announcements were made at school to alert students, notification was sent to students in their homeroom at school, two phone calls were made and a letter mailed to their residence, and computer labs were reserved for our exclusive use. The computer survey and the Mock Job Interview were scheduled together on a Saturday at their school and \$50 was provided immediately upon completion.

⁴ Participants also completed the computer survey at the end of the fall semester. Results from this survey were not meaningfully different than from the posttest survey and did not change our understanding of the treatment outcome. Therefore we are not presenting findings from this intermediate assessment.

Measures

We made minor modifications to many of the instruments. Specifically, some instruments response choices were rescaled and questions were reworded in order to make them more understandable for high school students. For some measures, we combined instruments and eliminated items to address our substantive interest and limit the length of the survey. We conducted extensive piloting with ASM apprentices to make sure the web-based computer survey was understandable. The psychometric properties (factor analysis, Cronbach's alpha) of the final instruments met accepted standards.

Background and Control Variables

Standard demographic information was obtained from youth report (computer survey), including gender, race/ethnicity, age, grade in school, receipt of free/reduced price lunch, and mother's educational attainment.⁵

We considered competing hypotheses that posit that non-ASM experiences account for any differential outcome. In particular, it may be that school extracurricular or paid employment experiences provide comparable or superior benefits to ASM, or that ASM gains accrue only to those apprentices who currently have, or have had in the past, paid employment experiences. Accordingly, we assessed participation in extracurricular activities at school or in community programs. Participants were asked to rate the amount of time they spent in five extracurricular activities (sports teams, performance programs, service activities, community programs, or academic/leadership clubs) over the course of the previous year. Work experience was assessed in detail, including type of work, training provided, amount of time at work, and hourly wage.

From archival data, we obtained the two most recent standardized achievement test scores (for most youth, their 7th and 8th grade scores). We created a mean of reading and math stanine scores across both years, and used the mean combined score as a control measure of academic ability.

Implementation/Process Variables

⁵ Receipt of free/reduced price lunch and mother's educational attainment were combined to make a single composite measure of SES (adapted from Ensminger et al., 2000).

Youth Program Quality Assessment. The High/Scope (2005) observational measure (YPQA) taps several program domains relevant to ASM: Safe Environment, Supportive Environment, Interaction, and Engagement. All of the scales, with the exception of safety (and we used only the psychological/emotional safety subscale), have good internal reliability. Because there are normative data available from a growing number of after-school programs, the YPQA enables us to determine where these ASM apprenticeships rank within the wider universe of after-school programs.

Additional observational data. Members of the research team observed each apprenticeship weekly and wrote detailed field notes. These data allowed us to assess program fidelity. The focus of this report is the quantitative findings from the evaluation; findings from qualitative observations are forthcoming in other papers.

Design features. We included a 7-item questionnaire to assess relatively objective design features of either the apprenticeship (for the apprenticeship youth) or the most time-intensive extracurricular activity (control youth). Items were chosen to assess key features (core elements) of ASM program objectives, developed in response to discussions with ASM. We conducted extensive piloting with ASM apprentices to make sure item and scale metric wording was understandable and salient. The measure is included in Appendix A. We report findings for individual items as well as the total score (a mean of all 7 items), which had an internal reliability $\alpha = .76$.

Social climate. Youth completed items from four of Moos' (1974) social climate measures: Involvement (from the Work Environment Scale), taps the extent to which apprentices put a lot of effort into their projects; Task Orientation (Work Environment Scale), measures the focus on getting work done; Cohesion (Group Environment Scale), assesses belongingness and group spirit; and Leader Support (Group Environment Scale), taps the supportiveness of the adult leader. The instrument instructions were reworded so that youth in the treatment were asked to rate their ASM apprenticeship and control group students were asked to rate their experience in their most time-intensive extracurricular activity. Items were also rescaled so that instead of true/false, participants were asked to rate the degree to which they agreed or disagreed with statements such as "The activities are really challenging," "People put

quite a lot of effort into what they do,” “There is a feeling of togetherness in this group,” and “The adult leader goes out of his/her way to help us in this activity.”

The total scale had reliability of $\alpha = .86$. Reliability for the Involvement subscale was very low ($\alpha = .34$) and therefore we do not report this subscale individually; items from this scale were included in calculating the total mean score. The other subscale reliabilities were acceptable (Task Orientation $\alpha = .60$; Leader Support $\alpha = .73$; Cohesion $\alpha = .77$) and are reported in the implementation section.

Dependent/Outcome Variables

Positive Youth Development

Self-efficacy. We acknowledge that there is a debate whether self-efficacy is task specific. We use a measure of global self-efficacy for several reasons. Firstly, the field is interested in whether after-school provides experiences that increase participants’ global self-efficacy, which would be tapped in multiple settings and not limited to the apprenticeship experience. We were repeatedly told that ASM at its best was able to “turn lives around,” which suggests a global rather than highly circumscribed impact. Certainly, most of the rhetoric about the benefits of after-school, and frequently cited qualitative studies (e.g., McLaughlin, Irby, & Langman, 1994), refer to a broad rather than narrow impact. Practically, in order to have task-specific measures of self-efficacy, we would need to assess efficacy related not only to each of the 13 different apprenticeships, but also each of the myriad extracurricular activities engaged in by the control group. It was not realistic to have available and get good data from such a multitude of instruments, which would also need to be equivalent and each psychometrically sound.

We used Sherer’s Self Efficacy Scale (1982). Sherer’s scale taps a number of relevant domains, including behavior initiation, effort, and persistence. The original 17-item scale has been the most widely used measure of this construct (Chen, Gully, & Eden, 2001). Our final scale was comprised of 13 items that included: “When I set goals for myself, I rarely achieve them,” “When I have something unpleasant to do, I stick to it until I finish it,” and “I feel

insecure about my ability to do things.” Youth were asked to rate how much they agreed or disagreed with each of the statements. The scale had an internal reliability of $\alpha = .87$.

Self-regulation. The scale taps the extent to which students are able to manage their attention and emotions as they pursue goal-directed activities. The original 10-item scale was in German and the scale authors developed an English translation (Luszczunska et al., 2003); we concluded that additional language modifications were necessary for suitability to the US high school setting. Students were asked to rate the degree they agreed or disagreed with statements such as, “I can concentrate on one activity for a long time if necessary,” “I can control my thoughts so they don’t distract me from what I’m doing,” and “After an interruption, I am able to concentrate again right away.” The final scale had an internal reliability of $\alpha = .83$.

Occupational values. The measure is the occupational values scale from the University of Minnesota’s Youth Development Study (Mortimer, 2003). The 12-item scale measures how important both intrinsic and extrinsic considerations are expected to be when an adolescent contemplates job seeking in his or her future. Items included “Good pay,” “Time off when I need it,” “A job that uses my skills and abilities,” and “A chance to be helpful to others or useful to society.” The total scale’s reliability was $\alpha = .83$.

Relationships with adult authority figures. This instrument is based on a measure developed by Barton J. Hirsch with David L. DuBois in recent after-school research funded by the WT Grant Foundation. The revised 10-item scale used in this study taps youth trust and caring in relation to adults and a belief that adults care about you as an individual and are interested in mentoring. Items include “Adults care about what happens to you,” “Adults are just interested in their job, not in you,” and “Adults are interested in helping you learn to be successful.” The internal reliability of the scale was $\alpha = .90$.

Interpersonal relationships. We also obtained self-report on skills that youth may have learned in ASM (or in their most time-intensive extracurricular activity for control youth) about how to work with others. We created an instrument using items from Youth Experience Survey (Hansen & Larson, 2002) supplemented with items from Mortimer’s Youth Development Survey. Items included “I learned to work with others on a team or group project,” “I learned

how my emotions and attitude affect others,” and “I learned to be able to give feedback.” Factor analysis confirmed measurement of a single construct tapping peer social skills; the final measure had a high internal reliability of $\alpha = .94$.

Marketable Job Skills

This domain was assessed with the Northwestern Mock Job Interview. Developed by the principal investigator (B. Hirsch) in collaboration with senior, Chicago-area human resource (HR) professionals, this structured interview is designed to sample job interview behavior by youth in a situation corresponding to an actual employment interview. We developed this instrument because the employment interview is the principal method by which employers make hiring decisions. Even social scientists who prefer that employers base such decisions on other kinds of data recognize that employers rely heavily on job interviews. For instance, Rosenbaum (2001) clearly preferred that employment decisions be based on school performance and teacher recommendations; nonetheless, in a study of employers who hire entry-level workers, he admitted that “All fifty-one employers we interviewed lean heavily on the information that they can gather themselves by directly interviewing applicants; all of the employers we spoke with rely on interviews [p. 131]. ...In sum, interviews are the primary determinant of hiring. From interviews, employers believe that they can infer which applicants have the requisite attitude, [and] interpersonal skills... Employers believe that they can make broad inferences from interviews, even if they last only fifteen minutes” (p. 134). We ourselves are agnostic as to whether employer reliance on job interviews is desirable; but if that is what employers rely on, then for evaluation research with policy implications it is important to use this methodology to determine whether ASM participation results in superior job skills.

As part of the Mock Job Interview, youth were given descriptions of two hypothetical summer jobs designed to have broad appeal and asked to indicate for which they were applying. They then completed a mock job application form. Experienced HR professionals conducted the interviews. Interview questions asked about prior experiences that applied to the position and targeted a range of desired job dispositions and skills. These included reliability and responsibility, goal orientation, and effort, as well as the ability to work collaboratively in teams, solve problems, communicate, learn on the job, and relate to a supervisor. In accord with

standard practice in employment research interviews (Eder & Harris, 1999), the questions were a mix of those that directly address prior experience with those that call for responses to hypothetical situations. Interviewers were not informed whether any youth was in the intervention or control condition. After completion of the 15-20 minute interview (and after the youth had left the room), the HR professional made 13 ratings on 1-5 scales (ratings of 1, 3, and 5 were anchored by specific descriptors) corresponding to each interview question. They then rated 12 youth qualities (confidence, ability to communicate clearly and persuasively, maturity, body language, etc.) based on the entire interview. Finally, they made overall judgments about whether they would hire the youth for a summer job and whether they would hire him for a permanent entry-level position. Because these latter ratings are of special importance, each of the 5 points of the scale is anchored by descriptors that correspond to actual hiring recommendations made by personnel officers.

Prior to the actual interviews, we trained the interviewers on the questions and use of the rating scales. They also were required to complete ratings on videotaped pilot interviews until they achieved 80% agreement with criterion ratings made by two senior HR professionals. The final interview had an internal reliability of $\alpha = .91$. The 12 interview questions asked of the participant had an internal reliability of $\alpha = .76$ and the 13 interviewer ratings of desirable applicant qualities had an internal reliability of $\alpha = .91$.

Academic Outcomes

School performance. We obtained information on grades and attendance from school records. GPA was calculated using weighted grades from students' core academic courses. We also used the number of failed core courses as an outcome measure.

School attendance. We used the number of days students were absent for each semester (and the total year) as reported on student transcripts.

Attitudes about school. Attitudes about school are measured using the Identification with School Questionnaire (Voelkl, 1996). The total scale included 15 items tapping students' self-reports of 1) a feeling of belonging at school and 2) a belief that school contributes to one's future success. Items include "I feel that I am treated with respect at school," "I'd rather be

anywhere else than in school” and “Most of what I learn in school will be useful when I get a job.” The two subscales (Belonging and Extrinsic Value, respectively) were validated in confirmatory factor analysis (Voelkl, 1996) and had acceptable internal reliability in our sample (Whole scale $\alpha = .78$; Belonging $\alpha = .73$; Extrinsic value $\alpha = .63$).

Problem Behaviors

A 10-item youth self-report measure of problem behaviors was taken from the National Longitudinal Study of Adolescent Health (ADD HEALTH; Undry, 1998). The scale includes measures from ADD HEALTH In-Home Interview from Wave II: Section 23 on Contraception; Section 27 on Tobacco, Alcohol, & Drugs; Section 28 on Delinquency; and Section 29 on Fighting and Violence. A total score was computed which represents the mean score across all 10 items ($\alpha = .77$).

Data Analysis Overview

We could analyze much of the data using analysis of variance methods, but because there is likely to be some missing data leading to imbalance in the design and because some of our outcomes are best considered to be discrete rather than continuous, we analyzed the data using a hierarchical linear model (HLM) approach (see, e.g., Raudenbush & Bryk, 2002). We carried out a hierarchical linear model analysis for each of the continuous dependent variables separately.

We estimated a two-level hierarchical linear model where the level one model is a within-apprenticeship-choice model. The level 2 model is an across apprenticeship model that could include apprenticeship-specific treatment effects. Because the students who select the same apprenticeship are likely to be more similar than those who select different apprenticeships, the sample *is* clustered by apprenticeship choice. In experimental design terminology, the apprenticeship choices are blocks and individual students are the replications within blocks. Because individuals within blocks (apprenticeship choices) are randomly assigned to treatments, the design is a (generalized) randomized blocks design with apprenticeships crossed with treatments (see, e.g., Kirk, 1995). Because the purpose of this study is to evaluate the effects of the apprenticeships that are part of the study and not to generalize to a broader population of apprenticeships, we treat the apprenticeships as having fixed effects in our analysis.

The specific level one model for the continuous outcomes can be described as follows. Let Y_{ij} be the outcome score of the j^{th} student who selected the i^{th} apprenticeship

$$Y_{ij} = \beta_{0i} + \beta_{1i}PRETEST_{ij} + \beta_{2i}MALE_{ij} + \beta_{3i}BLACK_{ij} + \beta_{4i}SES_{ij} + \beta_{5i}AGE_{ij} \\ + \beta_{6i}WORKEXP_{ij} + \beta_{7i}EXTRACURRIC_{ij} + \beta_{8i}PRIOR_{ij} + \beta_{9i}TREATMENT_{ij} + e_{ij}$$

where $PRETEST_{ij}$ is a covariate matched to the outcome, $MALE_{ij}$ is a dummy variable for gender, $BLACK_{ij}$ is a dummy variable for African American, SES_{ij} is a composite socioeconomic status variable (created from self-report free/reduced price lunch qualification and mother's educational attainment), AGE_{ij} is a measure of student age in years and months, $WORKEXP_{ij}$ is a dummy for work experience as reported at the pretest survey, $EXTRACURRIC_{ij}$ is a mean score of the amount of time spent across five extracurricular activities as reported at the pretest survey, $PRIOR_{ij}$ is a dummy variable for prior participation in an ASM program, $TREATMENT_{ij}$ is a dummy variable for the treatment condition, and e_{ij} is a student-specific residual. In this model, β_{1i} , β_{2i} , β_{3i} , β_{4i} , β_{5i} , β_{6i} , β_{7i} , β_{8i} and β_{9i} are the effects of $PRETEST$, $MALE$, $BLACK$, SES , AGE , $WORKEXP$, $EXTRACURRIC$, $PRIOR$, and $TREATMENT$ in apprenticeship i .

For most of the outcome variables, the covariate we have labeled $PRETEST$ is the same variable as the outcome, measured prior to the beginning of the intervention. There are two exceptions: marketable job skills and academic outcomes. Marketable job skills were obtained from the Mock Job Interview conducted only at the posttest assessment, and therefore this data was not available as a pretest covariate. For academic performance, we controlled for prior academic ability, as measured by a composite measure averaging the two most recent reading and math standardized test stanine scores.

We treat apprenticeship selection as a fixed effect, so the specific level 2 model for the level 1 intercept is

$$\beta_{0i} = \pi_{00} + \eta_{0i},$$

where π_{00} is the control group average across apprenticeships and η_{0j} is an apprenticeship-specific residual. The level 2 model for the other covariates is

$$\beta_{1i} = \pi_{10},$$

$$\beta_{2i} = \pi_{20},$$

$$\beta_{3i} = \pi_{30},$$

$$\beta_{4i} = \pi_{40},$$

$$\beta_{5i} = \pi_{50},$$

$$\beta_{6i} = \pi_{60},$$

$$\beta_{7i} = \pi_{70},$$

$$\beta_{8i} = \pi_{80},$$

so that the effects of the covariates are constrained to be equal across groups. In the primary analysis, the level 2 model for the treatment effect is

$$\beta_{9i} = \pi_{90}$$

so that we will be estimating and testing the average treatment effect across apprenticeships (π_{90}) controlling for the covariates.

The main object of the primary analysis for each dependent variable is to estimate and test the statistical significance of the average apprenticeship effect π_{90} .

CHAPTER 3

IMPLEMENTATION

This report presents key quantitative implementation findings. We also conducted weekly qualitative observations for each of the 13 apprenticeships involved in the evaluation. In this chapter, we first describe observational ratings completed by research staff. The YPQA enables us to determine where these ASM apprenticeships rank within the wider universe of after-school programs. We also report findings from youth self-report ratings of their apprenticeship experience.

After we discovered that the overwhelming majority of control youth were involved in extracurricular activities, we changed our implementation assessment measures in Year 2 to include control group ratings of their most time intensive extracurricular activity. For both the Design Features and Moos Social Climate measures, we report group means, the effect size, and tests of statistical significance for treatment (treatment-on-the-treated) and control group differences. In the final section of the chapter, we expand our consideration of non-ASM activities by reporting additional data on non-ASM activities for both groups.

High/Scope

Each program was rated using three subscales⁶ of the High/Scope (2005) observational measure (YPQA): Supportive Environment, Interaction, and Engagement. Programs were observed once per week for the entire year (roughly 1/3 of program sessions). Our ratings on the YPQA were based on observer impressions over the course of the entire year.

The YPQA enabled us compare ASM apprenticeships to a larger sample of after-school programs. For this comparison, we used the second wave from the High/Scope validation study (Smith & Hoffman, 2005). Their sample includes observational ratings from 118 programs, which served a wide range of middle and high school students and spanned a variety of topics.

⁶ We did not include the Safe Environment subscale as the physical environment items on this measure were not applicable.

The majority of the programs in their sample took place at least once per week, were after-school, and in urban settings. **Table 3.1** lists the mean ratings for the High/Scope sample and for the 13 programs in our study. ASM programs in the study were rated similarly to the High/Scope sample with regard to Supportive Environment. However, the programs in our evaluation were given higher scores on the Interaction and Engagement subscales compared to the High/Scope sample.

Program Design Features

We included a self-report questionnaire to assess relatively objective design features of either the apprenticeship for the apprenticeship youth or the most time-intensive activity for the control youth. The seven items were chosen to assess core elements of ASM program objectives, developed in response to discussions with ASM. Participants were asked to rate the frequency with which each design feature was implemented in their apprenticeship or extracurricular activity.

As can be seen in **Table 3.2**, the treatment group had a higher total scale score (mean of all 7 items) on the design features measure than the control group ($p < .001$), with a treatment effect of nearly one standard deviation ($g = .89$). Although we are primarily interested in scale scores rather than scores on individual items, here (and elsewhere in this report) we also present findings on individual items when it appears that doing so would increase our understanding of significant scale differences. Furthermore, we consider this an archival document and prefer to present a more rather than less comprehensive report.

As can be seen in **Table 3.2**, the treatment group rated the ASM apprenticeships significantly higher on 6 of the 7 key design features than the control group rated their extracurricular activities. Apprentices in the treatment reported the instructors taught them a skill ($p < .001$), how to improve their skills ($p < .001$), and how the activities were related to a job ($p < .001$) significantly more often than the adult leader of activities rated by the control group. The apprentices also rated having a deadline ($p < .001$), being able to make choices related to the activity ($p < .001$), and working on a project that was used or viewed by others ($p < .01$) as happening more often than what the control group reported for their activities. There was

no significant difference in the amount of time apprenticeship and control groups reported that they worked in groups with other teens in the activity.

Social Climate

Youth completed four of Moos' (1974) social climate measures: Involvement, Task Orientation, Cohesion, and Leader Support. As stated in the method section, the Involvement subscale had very low reliability and was not used as an independent subscale; however, items from this scale were used to calculate the Social Climate mean score. Similar to the design features instrument, youth in the treatment were asked to rate their ASM apprenticeship and control group youth were asked to rate their most time-intensive extracurricular activity. Participants were asked to rate the degree to which they agreed or disagreed with statements describing their activity; higher scores are more positive ratings on each scale.

Apprenticeship and control group means are reported in **Table 3.3**. Apprenticeships were not rated significantly different than control group extracurricular activities on the total scale, nor were any of the subscale differences significant.

Thus, on the quantitative implementation measures, ASM apprenticeships and the control group extracurricular activities did not differ in terms of their social climate, indicating that not only were control group youth involved in alternative activities, but that those activities were comparable in terms of important social-psychological dimensions. Where ASM and control group activities did differ was in terms of the design features that were fundamental to ASM.

Implementation and Outcome Measure Correlations

In order to further examine the possible effects of program implementation, we examined the correlation between dependent variables and 1) the Design Features total score and 2) Social Climate total score. Results are reported in **Table 3.4**.

Ratings of program design features were correlated with participant ratings of several measures of positive youth development (self-efficacy, self-regulation, adult relationships, and interpersonal learning) and to the school identification measure. Although significant statistically, the size of these correlations are modest. There was no significant association with

any measure of academic performance or attendance, nor with the problem behavior scale or interviewer ratings of marketable job skills.

The Social Climate ratings were more strongly correlated with outcomes.⁷ In addition, the social climate ratings were significantly correlated with outcomes that were not assessed via youth self-report, which is impressive as it is harder to find associations across methods. Social climate was highly correlated with participants' self report ratings on all measures of positive youth development. Participant ratings of social climate were also significantly correlated with interviewer ratings of marketable job skills, an objective measure of academic performance (weighted GPA), and the school identification measure.

Qualitative Observations

Our qualitative research did not reveal any major implementation problems aside from the high attrition rate. Consistent with the program design quantitative data, all programs were operated according to the ASM model; in particular, there was a strong emphasis on developing skills in all of the apprenticeships. There were no instances when programs did not meet for extended periods or did not obtain needed equipment within a reasonable time period. In two programs, a co-instructor was absent for significant stretches in the second semester for personal reasons.

There was an issue involving treatment integrity with one apprenticeship. The sports apprenticeship was presented as an opportunity to learn how to coach younger children. The apprenticeship taught skills in three different sports as well as games that would engage children. A number of students signed up for the apprenticeship because they were attracted to the opportunity to be involved with children. However, although the apprenticeship was designed to include visits by children during which apprentices could practice their coaching skills, no visits by children ever took place. In addition, in two of the sports, the co-instructor had either modest or no knowledge and skills in that particular activity, although the program was designed to rotate in co-instructors who had expertise in that particular sport.

⁷ Using Fisher r-to-z comparisons, positive youth development and school identification correlations with social climate ratings were significantly greater than correlations with design feature ratings.

While the apprenticeships included in the evaluation were not the very best exclusively, they were considered among the better ones in ASM at the time of their selection. Not surprisingly, we found several of them to be exemplary. To operationalize this a bit, had we been asked by editors of a volume that was going to feature exemplary after-school programs whether we could present any of the programs in our study, we would have responded affirmatively, and can think of five apprenticeships that would have been on our short list to write about. In all five instances, the apprenticeships produced interesting, original products and the instructors did a good job of creating a positive, effective social climate. For example, students in the Culinary apprenticeship focused intensely on producing food and food presentations of very high quality, and those in the Web Design apprenticeship produced attractive web sites for imaginary stores that sold products that were of personal interest to the apprentice who designed the site.

However, of the 13 apprenticeships, three did not meet our expectations in terms of some important dimension of program quality. One of the instructors had such poor communication skills that the students were often confused or treated in a hostile manner. Another instructor had such weak expectations for quality that many of the student products resembled those that might be found in an elementary school arts-and-crafts program. A third instructor never learned the names of most of her students, was unfocused, and all but three youth eventually dropped out.

Our observations are consistent with the social climate data in that apprenticeships varied considerably on those variables. Although there were programs that created very strong social climates, ASM did not strike us, overall, as superior in this domain. Over the course of many sessions, it was not unusual for there to be issues with engagement and time spent on task, with frequent episodes of socializing and web surfing.

Participation in Alternative Activities

As it became clear that the members of the control group participated in alternative activities, we sought to understand better the nature of those activities. We also sought to clarify how the control group's experiences might contrast to those of ASM youth, given that some ASM youth also participated in additional activities beyond ASM. Some of these analyses, which are based on data obtained in our spring assessment, will also help us to interpret impact findings on problem behavior that are presented later.

Although 91% of control youth participated in some structured after-school activity, it also was the case that 87% of ASM youth participated in an extra-curricular activity in addition to ASM. There were no significant between-group difference in participation rates.

It does not appear that the control group youth spent as much time overall in their activities as ASM youth spent in the combination of their ASM apprenticeship and other activities. When we looked at time investment across five types of non-ASM activities—sports, music/performance, service, community, and academic—it was only in sports that the control group spent significantly more time than did ASM youth (Kruskal Wallis $\chi^2 = 6.55, p = .01$). However, even here the difference between control youth (26%) and ASM youth (20%) who participated more than once/week (the most intensive involvement on our scale) was not substantial. When one adds together time ASM youth spent in ASM (three times per week) and outside of ASM, it appears likely that they spent more time in structured after-school activities than did the control group, though our measures do not allow us to test this directly.

In terms of paid employment outside of ASM, control youth were significantly more likely to report a paid job (23%) than ASM youth (14%; $\chi^2 = 5.37, p < .03$). However, there were no significant differences between the groups (for non-ASM positions) in terms of hourly wages, hours worked per week, or weekly earnings. Thus, if one adds in the additional \$45/week stipend that ASM youth received for their apprenticeships, it seems likely that ASM youth received more earnings than did control youth, although we do not have definitive data about this.

CHAPTER 4

IMPACT FINDINGS: INTENT-TO-TREAT

We first completed analyses examining the effect of treatment assignment using the entire participant sample. For this set of analyses, the treatment group was defined as having been assigned to the apprenticeship (which included those who received treatment and those who dropped out). The great majority of control group youth were involved in an alternative activity.

Each domain was measured using multiple dependent variables. We estimated HLM models for all dependent measures separately. Each model included key demographic variables and, when available, a matched pretest measure covariate. We report the t-ratio and statistical significance of the average apprenticeship effect. Standardized mean difference effect sizes were calculated using Hedges *g*.

Positive Youth Development Outcomes

Positive youth development outcome measures were obtained using a self-report computer survey. Participants were asked to complete ratings for self-efficacy, self-regulation, adult relationships, and occupational values before treatment assignment (pretest) and again at the posttest. Interpersonal relationships were assessed in relation to the apprenticeship (or controls' most time-intensive extracurricular activity) and therefore only assessed at the posttest survey. Students rated all measures in this domain on a scale from 1 (strongly disagree) to 5 (strongly agree); a higher mean score signifies greater self-perceived positive development. Pretest and posttest means for apprentice and control groups and treatment effect sizes are reported in **Table 4.1**. The last two columns list the t-ratio and corresponding significance test for the treatment effect from the HLM analyses.

There was a significant treatment effect for self-regulation. Both the intent-to-treat and control group declined in self-reported ratings of self-regulation from pre to posttest; however, treatment group means declined significantly less than the control group ($t\text{-ratio} = 2.14, p = .03$),

indicating a preventive impact. This translates to a positive treatment effect of .18 of a standard deviation after controlling for the pretest measure. Treatment effects for other four positive youth development outcomes failed to reach significance.⁸

Marketable Job Skills

Marketable job skills were assessed using the Mock Job Interview completed at the same time as the posttest survey. Due to the logistics and expense involved, the interview was conducted only once; therefore a pretest measure was not included as a covariate in the final HLM model.

The interview consisted of 13 questions. HR interviewers rated participant responses on a scale of 1-5 (higher scores indicate better answers). After the interview, HR interviewers also rated the applicant on 12 work-desirable qualities (1-5, higher scores indicated desirable ratings). Final hiring determination was also made on a 5-point scale (1 = definitely not hire; 5 = definitely hire) where HR interviewers were asked whether they would hire the participant for a summer job and whether they would hire them for an entry-level, full-time position. Hiring for the entry-level position was also scored dichotomously, where a score of 4 or 5 was considered 'hired' and a score of 1-3 was considered 'not hired'.

As can be seen in the first row in **Table 4.2** students assigned to treatment were not more likely to be hired for an entry-level position than the control group ($\chi^2 = .28, p = .60$). **Table 4.2** also contains treatment and control group means and standard deviations for hiring outcomes and HR ratings of interview questions and applicant qualities. There were no differences between the intent-to-treat and control groups on composite indices measuring HR ratings for the interview questions (Q1A through Q13A) or ratings of applicant qualities (Q1B through Q12B). For archival purposes, we also present scores on each of the items that make up the composite indices.

The lack of a more pronounced benefit to ASM participants in this area is consistent with our qualitative observations. Although some instructors explicitly referred to how hard skills (technical skills required in a particular occupation) or soft skills (generic work skills useful in a

⁸ External and internal occupational values subscales were also non-significant.

wide variety of fields) obtained in the apprenticeship transferred to actual employment situations, most did not. Moreover, the most frequent reference to transferability was a negative one, in which youth were scolded for their behavior and warned that, “You won’t be able to get away with that on a real job!” Identification of positive soft skills that were being learned were rare and there were fewer instances than we expected in which marketable hard skills were discussed in terms of actual employment opportunities. For example, in one otherwise excellent apprenticeship, which focused on web design, the only instance in which we observed a reference to the fact that youth could get paid for their design skills came during a presentation by a guest instructor. It therefore was not surprising when the human resource interviewers, during debriefing sessions, informed us that many of the young people had experiences and skills that employers would value, but that the youth themselves often had no idea that this was the case and did not successfully communicate those credentials in the Mock Job Interview.

Academic Outcomes

Academic performance. Academic performance outcomes were assessed using weighted grade point average and number of core courses failed for each semester. For these analyses, student standardized test scores were used as pretest control. As can be seen in **Table 4.3** there were no significant treatment effects for grade point average or for number of courses failed.⁹

School attendance. Because prior year attendance was not available for many of the students in the study, we estimated two separate HLM models: one using the subsample of participants with available prior year attendance records and a second for the entire sample using standardized test scores as an additional covariate.

We were able to obtain prior year attendance only for students who attended the same school the prior year (i.e., non-transfer students in 10th-12th grade during the evaluation year). For this subsample, we used prior year attendance as a pretest measure. **Table 4.4** reports the

⁹ We also considered the fact that some students are more likely to take more core courses and calculated a percentage of core courses failed as an outcome. This produced essentially the same result; there were no marginal or significant treatment effects.

number of students included in the analyses along with the mean group absences for the prior and current year. Overall, there were no significant treatment effects for school attendance.

For the whole sample (**Table 4.5**), we tested for treatment effects using a model with standardized test scores as an additional covariate, hypothesizing that students who have higher academic ability may be more likely to attend school. There were no significant differences between the intent-to-treat and control groups for school attendance.¹⁰

Attitudes about school. Students' perception of school was assessed at the posttest survey using the Identification with School Questionnaire, which asks students to rate their sense of belonging at school as well as their perception that school has value for their future pursuits. Higher mean scores indicate a more positive opinion of school (belongingness and value). Results are reported in **Table 4.6**. Because there was no pretest assessment, we first ran the model using only demographic controls (labeled Model 1). In this model, youth assigned to treatment had a higher identification with school compared to the control group (t-ratio = 2.22, $p = .03$), yielding an effect size of $g = .23$. Specifically, the intent-to-treat group reported significantly more value of school (t-ratio = 1.97, $p = .05$). Their scores on sense of belonging (t-ratio = 1.82, $p = .07$) were only marginally better compared to control group students.

Given that students with greater academic ability may be more likely to see value in school and to feel welcome, we estimated a second model controlling for student standardized test scores. Corresponding treatment T-ratio and p-values are labeled as Model 2 (see **Table 4.6**). In this model, students assigned to treatment had a marginally higher mean score for the overall Identification with School Questionnaire (t-ratio = 1.89, $p = .06$) compared to the control group. The difference with respect to extrinsic value (t-ratio = 1.81, $p = .07$) was also only marginally significant. There was no significant or marginally significant difference in sense of belonging (t-ratio = 1.44, $p = .15$).

¹⁰ It should be noted that for the longitudinal measures of academic performance (i.e., grades and attendance), there were declines for both groups from fall to spring.

Problem Behaviors

Participants reported at the pretest and posttest survey the frequency with which they engaged in 10 problem behaviors over the previous two months. A composite measure was calculated that represented the mean score averaged across all 10 items. HLM analyses estimated impacts controlling for background variables and corresponding pretest measure.

As seen in **Table 4.7**, there was a significant group difference (t-ratio = -1.95, $p = .05$) favoring youth assigned to ASM apprenticeships. Both groups reported a slight rise in problem behavior, but the rise was less for those assigned to ASM, suggesting a preventive impact. This impact seems driven by the crime items: youth assigned to ASM were significantly less likely to report selling drugs or participating in gang activity and marginally less likely to report stealing items worth more than \$50.

Power Analysis

To better interpret the treatment effects that were not found to be statistically significant, we carried out retrospective power analysis of all of our impact analyses (using the sample sizes we actually obtained). We computed statistical power to detect an effect size of 0.30 (0.31 was the mean effect size for studies that obtained significant outcome effects in the Durlak et al., 2010, meta-analysis). In the case of positive youth development, problem behavior, and academic outcomes, we assumed a covariate outcome multiple correlation of $R^2 = .50$ (a value which is conservative given our observed correlations). In the case of marketable job skills, we assumed no covariate, corresponding to the impact analyses we actually carried out. The analyses of impacts on positive youth development and problem behavior outcomes had a power of 95% to detect a true effect size of $d = .30$ (.30 standard deviations). The analysis of impacts on academic outcomes had a power of 97% to detect a true effect size of $d = .30$ (.30 standard deviations). The analysis of impacts on job skills had a power of 88% to detect a true effect size of $d = .30$ (.30 standard deviations). This suggests that there is reason to believe that the outcomes for which we did not find statistically significant effects had true effects less than 0.30 in standard deviation units.

One might argue that one would expect a somewhat lower effect size in an intent-to-treat analysis than in a treatment-on-the-treated analysis (although we found the differences to be minimal in our impact analyses). Making the same assumptions as above, the analyses of impacts on positive youth development and problem behavior outcomes had a power of 77% to detect a true effect size of $d = .20$ (.20 standard deviations). The analysis of impacts on academic outcomes had a power of 82% to detect a true effect size of $d = .20$ (.20 standard deviations). The analysis of impacts on job skills had a power of 48% to detect a true effect size of $d = .20$ (.20 standard deviations).

Heterogeneity of Treatment Effects

The study was planned and has adequate power to test the treatment effects that are constant across apprenticeships. It was not designed to have adequate power in the context of heterogeneous treatment effects. However we carried out a set of sensitivity analyses to explore the possible heterogeneity of treatment effects. These analyses were performed by changing the level 2 model for the treatment effect to permit it to vary across apprenticeships. Thus the level 2 model for β_{i9} (the treatment effect in the i^{th} apprenticeship) became

$$\beta_{i9} = \pi_{09} + \eta_{i9},$$

where π_{09} is the average treatment effect, η_{i9} is the deviation of the treatment effect in the i^{th} apprenticeship from that of the average treatment effect. We tested for heterogeneity of treatment effects by testing whether the variance (component) of the η_{i9} s is zero using a likelihood ratio test.

We carried out 27 tests for heterogeneity of treatment effects for outcomes in all four categories (positive youth development, marketable job skills, academic performance, and problem behavior). Tests were conducted for all scale scores, each of the academic outcomes, and for any individual item that had been found significant.

The heterogeneity of treatment effects was statistically significant at the 5% level in only 2 (7%) of these analyses. Because these are post hoc analyses and the rate of statistically significant results is indistinguishable from the rate that would be expected by chance, we interpret them as providing no evidence of heterogeneity of treatment effects across

apprenticeships. However, these analyses of heterogeneity have very low statistical power and thus provide only very weak evidence that there is no heterogeneity.

CHAPTER 5

IMPACT FINDINGS: TREATMENT-ON-THE-TREATED

An important question addressed by this evaluation is whether participation in ASM apprenticeships results in gains in positive youth development, marketable job skills, and academic performance, as well as decreases in problem behavior, that are above and beyond those obtained by the control group (most of whom are involved in alternative activities). In this chapter, we report the impact findings for those who *received* treatment using ASM's attendance standard.

As with the intend-to-treat analyses, we ran HLM analyses for all dependent variable measures separately. Each model included key demographic variables and, when available, a matched pretest measure covariate. We report the t-ratio and statistical significance of the average apprenticeship effect π_{90} . Standardized mean difference effects sizes were calculated using Hedges g .

We defined treatment participation using ASM's own standard for treatment dosage, which was a 73% attendance rate (see Chapter 2 for more details on this definition). Those who were assigned to treatment and were present for 73% of the apprenticeship sessions during the evaluation year were considered to have received treatment. The following reports impact findings in each of the key outcome domains using this treatment definition.

Positive Youth Development Outcomes

Positive youth development outcome measures were obtained using a self-report computer survey. Results are reported in **Table 5.1**.

The only significant treatment effect was for self-regulation, which declined for both treatment and control groups, but declined significantly less for the treatment group. The difference in relative decline translates to a positive treatment effect of $g = .19$ of a standard deviation after controlling for the pretest measure. The treatment effect estimated from the HLM analysis (controlling for background and pretest measures) was significant (t-ratio = 2.15, $p =$

.03). Treatment effects for the other four positive youth development outcomes failed to reach significance.¹¹

Marketable Job Skills

HR interviewers rated participant responses to 13 interview questions and rated the applicant on 12 work-desirable qualities. All ratings were on a scale of 1-5 (higher scores indicate better answers/more desirable qualities). Final hiring determination was also made on a 5-point scale (1 = definitely not hire; 5 = definitely hire) where HR interviewers were asked whether they would hire the participant for a summer job and whether they would hire him/her for an entry-level, full time position. The first row in **Table 5.2** reports the percentage of students hired from the apprenticeship and control groups. A Chi-square for group differences was not significant ($\chi^2 = .11, p = .75$).

Table 5.2 also contains treatment and control group means and standard deviations for hiring outcomes and HR ratings of interview questions and applicant qualities. Apprentices received significantly higher HR ratings on a summary index of the interview questions compared to the control group (t-ratio = 1.97, $p = .05$). HR professionals also rated the treatment group marginally higher on a summary index of work-related applicant qualities (t-ratio = 1.76; $p = .08$). There were few significant differences between the apprenticeship and control group on individual responses (see **Table 5.2**).

Academic Outcomes

Academic performance. Academic performance outcomes were assessed using weighted grade point average and number of core courses failed for each semester. For these analyses, student standardized test scores were used as a pretest control. As can be seen in **Table 5.3** there were no significant treatment effects for grade point average or for number of courses failed.¹²

¹¹ We also assessed the external and internal occupational values subscales as additional dependent variables. Results were also non-significant.

¹² We also considered the fact that some students are more likely to take more core courses and calculated a percentage of core courses failed as an outcome. This produced essentially the same result; there were no marginal or significant treatment effects.

School attendance. Prior year attendance was not available for many of the students in the study; specifically, were able to obtain prior year attendance only for students who attended the same school the prior year (i.e., non-transfer students in 10th-12th grade during the evaluation year). Therefore, we first analyzed treatment effects controlling for prior year attendance with this subsample of students. **Table 5.4** reports the number of students included in the analyses along with the mean group absences for the prior and current year. As seen in the table, there were no significant treatment effects.

We also completed a second analysis using the entire sample of students. We reran the HLM analysis using standardized test scores as an additional covariate, hypothesizing that students who have higher academic ability may be more likely to attend school (**Table 5.5**). Overall, students in the apprenticeship were absent fewer days in the fall semester compared with the control group. This difference translates into a treatment effect of $g = -.22$ of a standard deviation. The treatment effect was significant for the fall semester ($t\text{-ratio} = -2.02, p = .04$). There was no significant difference between the apprenticeship and control groups for spring attendance, and the treatment effect for the whole year was marginal ($t\text{-ratio} = -1.68, p = .09$).

Attitudes about school. Students' perception of school was assessed at the posttest survey using the Identification with School Questionnaire, which asks students to rate their sense of belonging at school as well as their perception that school has value for their future pursuits. Higher mean scores indicate a more positive opinion of school (belongingness and value). Results are reported in **Table 5.6**. Because there was no pretest assessment, we first ran the model using only demographic controls (labeled Model 1). Overall, apprentices had a higher identification with school compared to the control group ($t\text{-ratio} = 2.25, p = .03$). Specifically, apprentices saw increased value in school compared to control group students ($t\text{-ratio} = 2.65, p = .01$), yielding a treatment effect size of $g = .33$.

Given that students with greater academic ability may be more likely to see the value of school and to feel welcome, we ran a second model controlling for student standardized test scores. Results are presented as Model 2 in **Table 5.6**. Similar to the original model, apprentices reported an overall stronger identification with school ($t\text{-ratio} = 2.30, p = .02$), specifically in relation to extrinsic value ($t\text{-ratio} = 2.74, p = .01$), compared to the control group.

Problem Behaviors

Participants reported at the pretest and posttest survey the frequency with which they engaged in 10 problem behaviors over the previous two months. HLM analyses estimated impacts controlling for background variables and corresponding pretest measure.

As can be seen in **Table 5.7**, there was no significant difference between the treatment and control groups on the composite index. As in the intent-to-treat analyses, apprentices reported selling drugs significantly less than controls, but there was no significant difference regarding participation in gang activity.

Power Analysis

To better interpret the treatment effects that were not found to be statistically significant, we carried out retrospective power analyses of all of our impact analyses (using the sample sizes we actually obtained). We computed statistical power to detect an effect size of 0.30 (again, 0.31 was the mean effect size for studies that obtained significant outcome effects in the Durlak et al., 2010, meta-analysis). In the case of positive youth development, problem behavior, and academic outcomes, we assumed a covariate outcome multiple correlation of $R^2 = .50$ (a value which is conservative given our observed correlations). In the case of marketable job skills, we assumed no covariate, corresponding to the impact analyses we actually carried out. The analyses of impacts on positive youth development and problem behavior outcomes had a power of 87% to detect a true effect size of $d = .30$ (.30 standard deviations). The analysis of impacts on academic outcomes had a power of 89% to detect a true effect size of $d = .30$ (.30 standard deviations). The analysis of impacts on job skills had a power of 76% to detect a true effect size of $d = .30$ (.30 standard deviations). This suggests that there is reason to believe that the outcomes for which we did not find statistically significant effects had true effects less than 0.30 in standard deviation units.

Heterogeneity of Treatment Effects

As in the intent-to-treat analyses, we conducted exploratory tests to consider the possible heterogeneity of treatment effects for the treatment-on-the-treated groups. Of 29 tests, only 1 (3%) was significant. Because these are post hoc analyses and the rate of statistically significant results is indistinguishable from the rate that would be expected by chance, we interpret them as providing no evidence of heterogeneity of treatment effects across apprenticeships. However, these analyses of heterogeneity have very low statistical power and thus provide only very weak evidence that there is no heterogeneity.¹³

¹³ We also conducted sensitivity analyses to determine whether treatment was beneficial to subgroups of youth: African American males (given policy importance and difficulty in engaging them in programs); students in their freshman year (who are at heightened risk for academic and psychosocial declines; see Simmons & Blyth, 1987); and students from low SES families. We also tested for the interaction of social climate ratings and treatment to address effects of implementation. HLM analyses found no significant interaction effects for African American males or socioeconomic status; there was one marginal effect for freshman in the treatment group; and no significant social climate by treatment effects. Thus, there was no indication that treatment was especially beneficial for any of these subgroups.

CHAPTER 6

DISCUSSION

After School Matters is one of the largest and best known after-school programs for high school students in the country. It is based on an apprenticeship-like model that involves young people in paid, project learning experiences in a wide array of fields, from the arts to technology. Prior studies suggested that ASM could lead to important gains, laying the foundation for this work.

This evaluation research had a number of strengths: it utilized a randomized controlled trial design, which provides the best way to separate program from selection effects; multiple outcomes were assessed across four domains, permitting a comprehensive look at possible effects; data were obtained via multiple methods from a variety of sources, minimizing the potential bias inherent in any one type of data; extensive process data were obtained from both the ASM and control groups to provide a better understanding of underlying processes; and statistical techniques were employed so that we could correct for clustered sampling. The selection of experienced ASM instructors enabled us to test whether ASM works when, on balance, it is implemented well.

In what follows, we first synthesize the outcome results to consider the effectiveness of these ASM apprenticeships compared to the experiences of youth in the control group, most of whom were involved in alternative activities. Are there benefits to ASM above and beyond those obtained from participation in alternative activities? We will present both positive and skeptical responses that can reasonably be made to this query. We then offer our own conclusions and turn toward the future, considering future directions for ASM as a program and for evaluation research on after-school programs.

Comparative Effectiveness of ASM apprenticeships

We first consider the results in terms of statistically significant differences on specific outcomes, and then turn to consider results in terms of average effects per outcome domain.

Specific outcomes: Self-regulation and problem behavior

The two scales on which significant effects were found are substantively important and their impacts can be reasonably explained by participation in ASM vs. control group activities. Self-regulation is an important personal skill that taps how well youth are able to pursue goal-directed activities through the management of their attention and emotions (Gestsdottir & Lerner, 2008; Karol, 1993; Luszczynska, et al., 2004; Owens & Schneider, 2005; Zimmerman, 1995a,b). ASM apprenticeships provide support for such skill development during most program sessions, particularly through their emphasis on a specific project. Project demands, underscored by deadlines, and the likelihood that the product would be viewed by outsiders, provide a structure that motivates one to maintain focus on the tasks at hand in spite of distractions.¹⁴ The findings on the design features measure indicate that ASM was more successful at providing this structure than were the alternative activities in which control youth participated.

The finding that ASM youth reported less of an increase in problem behavior over time, particularly in relation to criminal acts such as selling drugs and participation in gang activities, is the result that is most likely to resonate in the policy world. There are several possible explanations for this finding. First, even when accounting for participation in non-ASM extracurricular activities, it appears that ASM youth spent more time in structured, after-school activities than did control youth. The after-school hours, particularly 3-6 PM, are well known as the period when youth are most at risk to be the victims or perpetrators of crime (Snyder & Sickmund, 1999). Thus, ASM youth are more likely to be unavailable during the highest-risk period to engage in juvenile crime. Second, when combining their ASM stipend with non-ASM work wages, it appears that ASM youth had higher earnings than did control youth. Accordingly, they may have less need for the financial benefits that can accrue from selling drugs or participating in gang actions. Finally, ASM design features may have led apprentices to be more likely to believe that there was a place for them in the adult employed world, an important developmental concern for adolescents (Erikson, 1968). There was a very strong

¹⁴ In intensive qualitative research at Boys & Girls Clubs, we found that conditions at the site helped distract youth from what could often be overwhelming stress in their lives (Hirsch, Deutsch, & DuBois, 2011, especially Chapter 3 on “Pocahontas”).

between-group effect ($g = .91$) for the item on that measure that tapped whether the adult leader talked about how the activity was related to an adult job/career. This suggests that ASM youth may be more likely to have positive expectations of meaningful adult employment and consequently less likely to jeopardize their perceived future for the short-term benefit of criminal activity.

Averaging effects within outcome domain

It is not unusual for treatment effects to be averaged within a particular outcome domain in meta-analyses or in reports such as those conducted by the What Works Clearinghouse. Such averages reflect treatment effects on broader constructs than any single measure. As ASM posits effects in broad outcome domains, this approach is appropriate for this evaluation. Given the large number of non-significant effects, it is not surprising that the average treatment effects within domains are small (see **Table 6.1**).

How do these results compare to other evaluations of after-school programs in the literature? Are the ASM effects stronger or weaker than has typically been found? We shall compare the ASM findings to those reported in two meta-analyses, each of which has its strengths and limitations.¹⁵

The recent Durlak et al. (2010) meta-analysis of after-school programs that seek to promote personal and social skills provides the most comprehensive comparative data. Across all 68 studies in the review, there was an overall mean effect size of 0.22, which involved averaging all effects within a single study and then averaging across studies. Their outcome domains do not correspond exactly with those of this evaluation; their domain of child self-perceptions is most similar to our positive youth development, and they do not report an overall academic domain, instead presenting effects for school grades, school attendance, and school bonding (which in our study is the measure of school identification). In looking at **Table 6.1**, we can see that the ASM effects for positive youth development (child self-perceptions), school attendance,

¹⁵ We will use the ASM treatment-on-the-treated figures as these map more closely unto those reported in the Durlak et al. meta-analysis.

and problem behavior are less; the ASM effects are roughly equivalent for grades, but higher for school identification (bonding).

Durlak et al. go on to classify the after-school programs in their review in terms of whether they are SAFE (i.e., sequenced, active, focused, and explicit). Significant effects were found only for those programs that were classified as SAFE (i.e., positive ratings on all four SAFE dimensions). We asked Durlak whether ASM would have been included in his review given the nature of the program and whether ASM would be classified as SAFE. After reading a draft of this report, Durlak concluded that ASM would have been included in his review. In terms of whether ASM constituted a SAFE program, Durlak indicated that he would rate it as having sequenced and active learning strategies, but not focused and explicit interventions with respect to socio-emotional learning. Compared to SAFE programs, ASM had smaller effect sizes for self-perceptions, grades, school attendance and problem behavior; the effect sizes were equal for school identification (see **Table 6.1**). Compared to non-SAFE programs, ASM had stronger effects for school identification and grades; roughly equivalent effects for school attendance and problem behavior; and weaker effects for positive youth development.

There are several limitations of the Durlak et al. (2010) review for our purposes. Only 35% of the studies they reviewed involved random assignment. It is possible that the average effect sizes are over-estimates, as more rigorous, experimental studies may obtain weaker effects. In addition, only 9% of the studies involved programs serving HS youth. Finally, many of the studies did not determine whether members of the comparison group were participating in alternative, formal after-school activities, so it is unclear whether they constitute a no-treatment or an alternative treatment control group. One would expect stronger effects if treatment was being compared to a no-treatment control.

A meta-analysis by Zief, Lauver, and Maynard (2006) focused exclusively on experimental evaluations of after-school programs. The effect sizes found in these studies were, indeed, lower than those found in the Durlak et al. (2010) review. Unfortunately, only five studies were available, none of which focused on high school youth. Moreover, all of the programs they reviewed included an academic component, which ASM does not have.

*

How do we weight all of these considerations? Most of us probably would like from this study a clear, definitive conclusion as to ASM's effectiveness or, more precisely, its added value compared to the experiences of the control group. However, we believe that there are two contrasting interpretations of these findings, each of which is reasonable. In order to best appreciate what this study does and does not tell us, we thought it best to elaborate each of these perspectives before offering our own conclusions.

A positive view of the evaluation findings

This perspective emphasizes that significant positive effects on important outcomes were found. Although the treatment-on-the-treated comparisons make the most intuitive sense in the program world, the most convincing findings to hard-core methodologists will be those from the intent-to-treat analyses (ITT). Given the considerable attrition in the ASM sample, the youth who remained in the treatment group may differ in some important way from the original, randomly assigned group. Although we addressed this empirically and found no meaningful differences on the data available, there may be differences on other variables that we did not measure. It is not unreasonable to suspect that there are underlying differences that lead some youth to drop out of ASM and others to remain, and that those who remain are more likely to improve in any case. The ITT analyses examine the ASM and control groups as originally assigned and thus remove this source of doubt. So when significant findings are obtained from ITT analyses, as was the case for the self-regulation and problem behavior scales, this should be convincing to those who are hardest to convince. Indeed, the strongest findings came from the strongest methodology.

The two scales on which significant effects were found—self-regulation and problem behavior—are substantively important. The problem behavior impact in particular is likely to resonate in the policy world. Moreover, the credibility and meaningfulness of these two findings are enhanced because they can be linked to specific components of the ASM experience that differed from the experience of control youth in alternative activities.

ASM programs were found to have stronger design features than the extracurricular programs in which control youth participated. These design features—including teaching specific skills, youth having a choice in activities, getting feedback on how to improve—have considerable support in the educational literature (e.g., Edelson, 2001; Brown & Camione, 1996; Rogoff, 1990; Schank, 1995) and in the National Research Council (2002) report on youth programs. Correlational analyses supported their value, as those variables were significantly associated with several outcome variables in our dataset.

ASM was able to obtain positive results despite several factors that worked against doing so. For example, although the research design specified the selection of experienced instructors with high skills, not all of the instructors proved to be superior. In addition, several of the programs had low attendance. Both of these factors should have decreased the effectiveness of ASM, but significant effects were still obtained. It is possible that effects would have been less if one had selected a wider range of instructors, but that is speculation; the findings that are available indicate that there are positive effects.

Most initial randomized trials are efficacy studies in which programs receive a considerable amount of extra support, especially in ongoing consultation to improve implementation, which increases the likelihood of positive outcomes. However, very little extra support was provided to the ASM apprenticeships in the study; there were efforts to get them supplies in a timely manner, a feedback session for executives and regional directors regarding instructors' lack of focus on linkage between the apprenticeships and the work world, and a one-session workshop for apprentices on job interview skills. In terms of support, these efforts are more in line with those of an effectiveness trial under typical implementation conditions, rather than the generous extra resources provided in efficacy trials. It is harder to obtain significant effects in effectiveness trials.

Almost all control group youth were involved in an alternative extracurricular activity or in paid work. This exposure meant that we needed to consider the control group as receiving an alternative treatment rather than as a no-treatment control. Clearly it is more difficult to find a treatment effect when the comparison group is also receiving treatment. Moreover, given that the comparison group had equivalent results on the social climate measures, which tap important

program dimensions, it is reasonable to consider them a strong comparison group. This increased the difficulty of finding a significant ASM treatment effect.

Finally, no significant negative effects were found in which control youth performed better than ASM youth. Negative effects have been found for some after-school programs (e.g., James-Burdumy, Dynarski, & Deke, 2007).

In summary, the positive perspective on the ASM evaluation emphasizes the significant positive effects that were found despite a number of high bars to obtaining such results.

A skeptical view of the evaluation findings

The skeptical perspective emphasizes that weak effects were found under conditions in which stronger effects should have been obtained. The study purposefully employed the better ASM instructors, maximizing the potential for finding strong, significant effects. All of the instructors had at least one year of experience and almost all had several years of experience. By contrast, there was no effort to obtain the best instructors or group leaders among the extracurricular programs of the control group youth.

Despite these advantages, when significant findings were obtained, the effect sizes were not large. Although there are no universally agreed upon standards for interpreting the strength of effect sizes, the typical view would likely be that the size of the self-regulation and problem behavior effects are not strong. When one examines the average effect size per domain across all four domains, the ASM effects appear weak. Although ASM's most important outcome is positive youth development, the average effect size in that domain is close to zero. There are no significant effects at all for ASM's second most important outcome domain, marketable job skills.

In summary, this perspective emphasizes that only a few significant effects were found, effect sizes were typically small, and that testing a more representative sample of ASM instructors may well eliminate the few positive impacts that were found.

Conclusions and Future Directions

We believe that each of these perspectives makes reasonable points, but that it is unwise to view this as a situation in which one needs to choose only one or the other. Although our culture increasingly expects a simple thumbs-up or thumbs-down summary judgment, it is important to take a more complex, historical perspective that considers how social science and program practice can interact over time. Very few randomized evaluation studies have been done on after-school programs and to the best of our knowledge only one such study has been done with an exclusively high school sample. High quality evaluation studies can provide information that is crucial to program development for strengthening the ability of programs to produce strong effects. In turn, redeveloped programs need to be subjected to further evaluation, which can result in a cumulative process that greatly enhances program effectiveness. The after-school world is just at the beginning of this process.

Within this broader context, we consider the ASM impacts to be promising. Although it is frequently the case that no significant treatment effects are found in experimental outcome studies, in this research ASM did have a significant impact in areas that are important to adolescent development and to policy. Moreover, it demonstrated these impacts in relation to what was essentially an alternative treatment comparison group. Nonetheless, we consider the counter-arguments from the skeptical perspective to be serious and these force us to view the outcomes with some caution. The caution is with respect primarily to whether the average ASM apprenticeship (which we did not study) is likely to provide outcomes superior to what high school youth can obtain in alternative extra-curricular activities, after-school programs, and part-time jobs. The skeptical perspective may ultimately prove prescient with respect to what such an outcome study would reveal. In order to maximize the likelihood that the modal ASM apprenticeship proves superior in such a future outcome evaluation, it is important to improve the program model.

What, then, are some important lessons for the future that can be learned from this study?

For research

Researchers can help the field to progress in a number of ways. Elaborated change models are needed that are based both on theory and on accumulating empirical findings. Such models should focus not only on the outcomes, but the pathways or mechanisms through which different outcomes are achieved over time (and perhaps over a longer time frame than employed in this research). It is critical that potential effects be more clearly conceptualized and new measurement instruments developed. We did this for marketable job skills, but much work remains to be done with respect to positive youth development. Existing measures in that domain have been taken primarily from research on child and adolescent development or from child clinical studies, but there are large gaps between available instrumentation and some of the youth changes that we heard described by program staff. There is still uncertainty regarding exactly what and how outcomes should be measured within the still emerging area of positive youth development and more progress needs to be made in this area. This should not be surprising given that, by contrast, there has been much more thinking and research about youth problem behavior over many more years (indeed, decades).

We have acknowledged that there is a debate as to whether certain types of effects are domain-specific or global, which is relevant to the selection of measures. In this study, the variable most implicated in this debate was our global measure of self-efficacy. The domain specific argument is that gains in self-efficacy with respect to math, for example, do not automatically translate to self-efficacy gains in science, or to the academic domain more generally. Unfortunately, there are practical difficulties in conducting domain-specific assessment over all the domains we would have needed to assess; in this study, this would include not only the 13 apprenticeships, but also the myriad extracurricular activities and jobs of control group youth. We have also found it necessary to interview control youth regarding which activity they should be rating to obtain valid data (Mekinda & Hirsch, 2010). Beyond technical issues in methodology, it is important to consider that documenting gains in global self-efficacy may well have more impact on policy makers than gains in a highly circumscribed activity (especially in what may be considered esoteric arts activities). Thus, the incremental

cost and utility of domain-specific assessment needs to be considered and not just its theoretical rationale.

The extent of control group participation in alternative activities came as a surprise. ASM had indicated to us that few such activities were available to this population of Chicago youth. Several school principals seconded this notion to us in preliminary operational meetings. The empirical research literature suggests a higher participation rate than our local estimates. National estimates for participation in at least one structured activity hover between 70 and 80% (Bouffard, 2006; Feldman & Matjasko, 2005; Mahoney, Vandell, Simpkins, & Zarrett, 2009), though there are consistent findings that low-income youth are under-represented in organized activities (Bartko & Eccles, 2003; Feldman & Matjasko, 2005; McNeal, 1998; Pedersen & Seidman, 2005). It is very difficult to find a precise published figure for the participation rate for a sample that is demographically comparable to ASM youth. In the only instance that we were able to identify, Pederson and Seidman (2005) reported that between 38-42% of the urban, low-income HS youth in their study participated in an organized school-based activity and between 23-26% participated in such a non-school activity. They do not report an overall participation rate across both types of activity, though it clearly would be less than the participation rate for our control group (91%).¹⁶ Differences in sample selection were likely a major factor accounting for our higher participation rate. In the other studies cited, researchers sampled the school population at large, which likely included many students with no structured, after-school activity involvement. In this study, all of the control youth had already sought to enroll in an after-school activity (i.e., ASM). We know from prior quantitative research that many young people who engage in one activity participate in others as well (e.g., Feldman & Matjasko, 2007), so the control youth were students likely to seek out additional activities when ASM was not available. This would be consistent with findings from qualitative studies that depict a distinct culture of highly involved students who are well-networked, well-informed of after-school opportunities,

¹⁶ Pederson and Seidman (2005) also reported that between 49-55% played team sports, but the phrasing of the item made it likely that youth would be counted who engaged in pickup games in a gym or playground setting that did not include adult involvement. If a substantial proportion of this group participated in an organized sports team led by an adult, then the cumulative involvement of that sample would more closely approach our figure.

and encouraged by friends, teachers, and activity leaders to attend (Flores-Gonzalez, 2002, 2005; Quiroz, Gonzalez, & Frank, 1996).

For purposes of evaluation research, in order to best understand program effects and their magnitude, more effort needs to be invested in documenting and understanding the experience of control groups. If treatment outcomes are determined by comparison to the experiences of controls, then the control group experience needs sufficient attention to justify interpretations that can have a profound effect on programs and policy. Both quantitative and qualitative studies of control groups are needed. These studies need to go beyond documentation of participation rates to consider as well the quality of control group experiences.

Finally, there is the issue of cost effectiveness. We did not collect data on ASM operating costs or the costs associated with the activities of control group youth. The issue of cost needs to be considered together with benefits for policy purposes.

For programs

Our quantitative and qualitative findings indicate that ASM programs need strengthening. This should not be a surprise or a disappointment; indeed, this is what evaluation research needs to find if the field is to advance.

Our observations revealed that ASM youth need to be more fully engaged. Too much time was spent surfing the web or socializing with each other. Greater effort also needs to be put into building high quality products or performances. To realize the potential of the ASM design features, much more attention needs to be paid to engage youth throughout a session. These are really fundamental principles of youth programs, so we will not elaborate on them here. But that should not be taken as a reflection that they are unimportant. ASM needs to spend a lot more time getting the basics down a lot better. This is true not only for instructors, but also for their supervisors, who often did not seem to know how to improve the performance of seriously deficient instructors.

Although many apprentices, in normal ASM practice, remain enrolled for more than one semester, in our study many instructors had difficulty in developing experiences in the second semester that differed meaningfully from those in the first semester. Apprentices often did learn

to perform skills faster and better, but the importance of such gains were not always communicated by the instructors. Because of drop-outs, many apprenticeships enrolled new students in the second semester (none of those new enrollees became subjects in our study). We were surprised by how quickly they came up to speed with those who had been continuously enrolled; had new, more advanced skills been taught, presumably this would have happened less quickly, if at all. In our view, instructors need much more training and supervision in how to develop additional, sequenced experiences over time, in how to build, in effect, a curriculum that deepens over more than one semester.

ASM can learn some important implementation lessons from its own best practices. The culinary arts program was one of the most effective programs we found for engaging youth and producing high quality work. The instructors divided the youth into two teams, which competed with each other to produce the best menus, food taste, and presentation. The quality of their efforts was judged by a team of adult outsiders, which typically included school staff (teachers, security guards) and occasional others (such as the research team). A scoring system was devised by the instructors to rate each of several components. The winners were announced to a great hurrah and the losers needed to do all the clean-up. The friendly competition resulted in sustained hard work and attention to detail, more so than in any other apprenticeship. The mock interview data analysis indicated that the culinary students had the best effect size vs. their controls of any of the apprenticeships when ratings were analyzed on the original 1-5 scale (they were third best when analyzed in terms of simply hired vs. not hired). We presented these findings to ASM and drew up a scheme that could be used for judging the quality of web design products to illustrate how the basic model could be applied to vastly different content areas and apprenticeships. We will discuss the culinary apprenticeship in depth in a subsequent paper being developed for publication.

ASM needs to focus more on transferable skills. This was especially evident regarding marketable job skills, which ASM has always highlighted as one of its primary concerns. To us, marketable meant that someone would find those skills relevant and valuable to their employment context. Given that many of these apprenticeships focused on occupations in which the number of jobs were quite limited (especially true for arts apprenticeships), the broader

utility of the skills for different occupations needs to be considered. The human resource interviewers indicated that this is always a major focus in interviewing for entry-level jobs. However, ASM instructors rarely alluded to such transfer and many youth did not appreciate that through ASM they had developed knowledge and skills that would be valued by employers. Young people need to be told that and to be trained specifically on how to communicate that information successfully in employment (and internship) interviews. Our research group recently piloted an intervention along these lines with career academies in Chicago Public Schools, with promising results, and similar efforts could be made with after-school programs.

Similar efforts at transferring knowledge, attitudes, and skills could be made regarding positive youth development. This would draw on the domain-specific argument in the measurement debate alluded to earlier as an impetus for innovative program development. Youth who learn skills in self-regulation in ASM, for example, could be explicitly trained to apply those in other situations. They could learn how to apply a new sense of self-efficacy gained in a specific after-school activity to other activities both in and out of school. This new direction would be consistent with findings regarding SAFE interventions from the after-school meta-analysis (Durlak et al., 2010): the program should be focused on and clearly emphasize specific skills. It was clearly the SAFE programs that produced the best results. A promising—and critical—new direction for the future is to focus explicitly on transferable skills.

In terms of implementation, there are important questions regarding how well apprenticeship instructors, trained in a specific craft, would be interested and skilled in providing such transfer-focused instruction. It may be necessary for ASM, and similar programs, to provide in-house consultants who work with instructors or provide the specialized training themselves.

ASM promotes itself as fostering positive youth development and marketable job skills and thus it must take issues of transfer of learning very seriously and experiment with new methods of working to make successful transfer a reality.

Final thoughts

This evaluation assessed ASM apprenticeships at a given point in time, at a still early point in its history. Since that time, ASM has informed us that they have improved instructor training and made other changes to better serve youth. Thus, caution needs to be exercised regarding the applicability of the findings to current and future ASM program offerings and to ASM as an organization.

Our broadest suggestion, to incorporate a focus—and provide explicit training—on the transfer of skills, poses a considerable intellectual and practical challenge to the existing ASM model. We believe that core ASM design features can be maintained, particularly project-based learning, and that these provide a promising foundation to build upon. However, to implement the additional elements well will require significant change. Given that the weakest outcome findings (averaged by domain) were in ASM’s two highest priority areas—positive youth development and marketable job skills—serious reflection is clearly demanded. To best enable youth to achieve those key objectives, it seems important to rethink important elements of what happens in the apprenticeships.

APPENDIX A
DESIGN FEATURES INSTRUMENT

Items

1. How often did an adult leader teach you some skills while you were involved in this activity?
2. How often did you have a deadline to accomplish something in this activity?
3. How often did an adult leader tell you what you needed to do to improve your skills in this activity?'
4. How often did the adult leader of the activity talk about a job/career that you could do related to this activity?
5. How often did you do this activity working together with other teens in groups?
6. How often were youth allowed to make some of their own choices about any part of the activity?
7. How often did you do something in this activity that was supposed to be used or viewed by others?

Rating Scale

- 1) Never or almost never
- 2) Less than half the time we met for this activity
- 3) About half the time we met for this activity
- 4) Most of the time we met for this activity
- 5) Every time, or almost every time, we met for this activity

REFERENCES

- Allensworth, E. (2005). *Graduation and dropout trends in Chicago: A look at cohorts of students from 1991 through 2004*. Chicago: Consortium on Chicago School Research.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman and Company.
- Bandura, A. (1995). *Self-efficacy in changing societies*. Cambridge, UK: Cambridge University Press.
- Barron, B. J. S. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *The Journal of the Learning Sciences*, 7, 272-322.
- Bartko, W. T., & Eccles, J. S. (2003). Adolescent participation in structured and unstructured activities: A person-oriented analysis. *Journal of Youth and Adolescence*, 32(4), 233-241.
- Benson, P. (1997). *All kids are our kids: What communities must do to raise caring and responsible children and adolescents*. San Francisco: Jossey-Bass.
- Bouffard, S. M., Wimer, C., Caronongan, P., Little, P. M. D., Dearing, E., & Simpkins, S. D. (2006). Demographic differences in patterns of youth out-of-school time activity participation. *Journal of Youth Development*, 1(1).
- Bretherton, I., & Munholland, K. (1999). Internal working models in attachment relationships: A construct revisited. In J. Cassidy & P. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 89-111). New York: Guilford.
- Browne, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glasner (Eds.), *Innovations in learning: New environments for education* (pp. 289-325). Mahwah, NJ: Erlbaum.
- Carnegie Corporation of New York. (1992). *A matter of time: Risk and opportunity in the nonschool hours*. New York: Self.
- Carr, R., Wright, J., & Brody, C. (1996). Effects of high school work experience a decade later: Evidence from the National Longitudinal Survey. *Sociology of Education*, 69, 66-81.

- Chen, G., Gully, S., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*, 62-83.
- Connell, J., Gambone, M., & Smith, T. (2000). *Youth development in community settings: Challenges to our field and our approach*. Philadelphia: Public/Private Ventures.
- Corno, L. (1993). The best laid plans: Modern conceptions of volition and educational research. *Educational Researcher, 22*, 13-22.
- Corporate Voices for Working Families (2006). *After school for all: A call to action from the business community*. Washington, DC: Self.
- Cotterell, J. L. (1996). *Social networks and social influences in adolescence*. New York: Routledge.
- DuBois, D., & Karcher, M. (Eds.). (2005). *Handbook of youth mentoring*. Thousand Oaks, CA: Sage.
- Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology, 46*, 294-309 .
- Eccles, J. S., & Barber, B. L. (1999). Student council, volunteering, basketball, or marching band: What kind of extracurricular involvement matters? *Journal of Adolescent Research, 14*(10), 10-43.
- Edelson, D. C. (2001). Learning-for-use: A framework for the design of technology-supported inquiry activities. *Journal of Research in Science Teaching, 28*, 355-385.
- Eder, R., & Harris, M. (Eds.). (1999). *The employment interview handbook*. Thousand Oaks, CA: Sage.
- Ellwood, D. (1982). Teenage unemployment: Permanent scars or temporary blemishes? In R. Freeman & D. Wise (Eds.), *The youth labor market problem*. Chicago: University of Chicago Press.
- Ensminger, M. E., Forrest, C. B., Riley, A. W., Kang, M. S., Green, B. F., Starfield, B., & Ryan, S. A. (2000). The validity of measures of socioeconomic status of adolescents. *Journal of Adolescent Research, 15*, 392-419.
- Erikson, E. (1968). *Identity: Youth and crisis*. New York: W.W. Norton.

- Feldman, A. F., & Matjasko, J. L. (2005). The role of school-based extracurricular activities in adolescent development: A comprehensive review and future directions. *Review of Educational Research, 75*(2), 159-210.
- Feldman, A. F., & Matjasko, J. L. (2007). Profiles and portfolios of adolescent school-based extracurricular activity participation. *Journal of Adolescence, 30*, 313-332.
- Flores-Gonzalez, N. (2002). *School kids/street kids: Identity development in Latino students*. New York: Teachers College Press.
- Flores-Gonzalez, N. (2005). Popularity versus respect: School structure, peer groups and Latino academic achievement. *International Journal of Qualitative Studies in Education, 18*, 625-642.
- Gestsdottir, S., & Lerner, R. (2008). Positive development in adolescence: The development and role of intentional self-regulation. *Human Development, 51*, 202-224.
- Gottfredson, D. C., Cross, A. B., Wilson, D., Rorie, M., Connell, N. (2010). Effects of participation in after-school programs for middle school students: A randomized trial. *Journal of Research on Educational Effectiveness, 3*, 282-313.
- Gottfredson, D. C., Gerstenblith, S. A., Soule, D. A., Womer, S. C., & Lu, S. (2004). Do after school programs reduce delinquency? *Prevention Science, 5*, 253-266.
- Goerge, R., Cusick, G. R., Wasserman, M., & Gladden, R. M. (2007). *After-school programs and academic impact: A study of Chicago's After School Matters*. Chicago, IL: Chapin Hall Center for Children at the University of Chicago.
- Greenberger, E., & Steinberg, L. (1986). *When teenagers work: The psychological and social costs of adolescent employment*. New York: Basic Books.
- Hahn, A., Leavitt, T., & Aaron, P. (1994). *Evaluation of the Quantum Opportunities Program demonstration: Short-term impacts*. Washington, DC: Mathematica Policy Research.
- Halpern, R. (2006). After-School Matters in Chicago: Apprenticeship as a model for youth programming. *Youth and Society, 38*, 203-235.
- Halpern, R. (2009). *The means to grow up: Reinventing apprenticeship as a developmental support in adolescence*. New York: Routledge.

- Halpern, R. with F. Kimondo (2005, January). *A Qualitative Study of After School Matters (Interim Report)*. Chicago, IL: Erikson Institute.
- Hamilton, S. (1990). *Apprenticeship for adulthood: Preparing youth for the future*. New York: The Free Press.
- Hamilton, S. F., & Hamilton, M. A. (2004). Contexts for mentoring: Adolescent-adult relationships in workplaces and communities. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 395-428). Hoboken, NJ: John Wiley & Sons.
- Hansen, D. M., & Larson, R. (2002). *The Youth Experience Survey*. Unpublished manuscript, University of Illinois at Urbana-Champaign.
- Hansen, D. M., Larson, R. W., & Dworkin, J. B. (2003). What adolescents learn in organized youth activities: A survey of self-reported developmental experiences. *Journal of Research on Adolescence*, 13, 25-55.
- Hawkins, J. D., Catalano, R., & Associates. (1992). *Communities that care: Action for drug abuse*. San Francisco: Jossey-Bass.
- High/Scope Educational Research Foundation. (2005). *Youth program quality assessment*. Ypsilanti, MI: High/Scope Press.
- Hirsch, B. J. (2005). *A place to call home: After-school programs for urban youth*. Washington, DC: American Psychological Association & New York: Teachers College Press.
- Hirsch, B. J., Deutsch, N. L., & DuBois, D. L. (2011). *After-school centers and youth development: Case studies of success and failure*. New York: Cambridge University Press.
- Holzer, H. (1996). *What employers want: Job prospects for less-educated workers*. New York: Russell Sage Foundation.
- Huffcutt, A., Conway, J., Roth, P., & Stone, N. (2001). Identification and meta-analytic assessment of psychological constructs. *Journal of Applied Psychology*, 86, 897-913.
- Jackson, S. L., Krajcik, J., & Soloway, E. (1998). The design of guided learner-adaptable scaffolding in interactive learning environments. Proceedings of ACM CHI'98, *ACM Press*, pp. 187-194.
- Karoly, P. (1993). Mechanisms of self-regulation: A system view. *Annual Review of Psychology*, 44, 23-52.

- Kemple, J., & Willner, C. J. (2008). *Career academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood*. New York: Manpower Development Research Corporation.
- Kirk, R. (1995). *Experimental design*. Belmont, CA: Brooks Cole.
- Larson, R. (2007). From “I” to “we:” Development of the capacity for teamwork in youth programs. In R. Silbereisen, & R. Lerner (Eds.), *Approaches to Positive Youth Development*. Thousand Oaks, CA: Sage.
- Larson, R., Jarrett, R., Hansen, D., Pearce, N., Sullivan, P., Walker, K., Watkins, N., & Wood, D. (2004). Organized youth activities as contexts for positive development. In A. Linley & S. Joseph (Eds.), *International handbook of positive psychology in practice: From research to application*. New York: Wiley.
- Lerner, R. M. (2004). *Liberty: Thriving and civic engagement among America’s youth*. Thousand Oaks, CA: Sage.
- Luszczynska, A., Diehl, M., Gutierrez-Dona, B., Kuusinen, P., & Schwarzer, R. (2004). Measuring one component of dispositional self-regulation: Attention control in goal pursuit. *Personality and Individual Differences, 37*, 555-566.
- Mahoney, J. L., Larson, R. W., & Eccles, J. S. (2005). *Organized activities as contexts of development: Extracurricular activities, after-school, and community programs*. Mahwah, NJ: Erlbaum.
- Mahoney, J. L., Vandell, D. L., Simpkins, S., & Zarrett, N. (2009). Adolescent out-of-school activities. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3 ed., Vol. 2: Contextual influences on adolescent development, pp. 228-269). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Maxfield, M., Schirm, A., & Rodriguez-Planas, N. (2003). *The Quantum Opportunities Program demonstration: Implementation and short-term impacts*. Washington, DC: Mathematica Policy Research.
- Maxfield, M., Castner, L., Maralani, V., & Vencill, M. (2003). *The Quantum Opportunities Program demonstration: Implementation findings*. Washington, DC: Mathematica Policy Research.

- McLaughlin, M., Irby, M. & Langman, J. (1994). *Urban sanctuaries: Neighborhood organizations in the lives and futures of inner-city youth*. San Francisco, CA: Jossey-Bass.
- McNeal, R. B. (1998). High school extracurricular activities: Closed structures and stratifying patterns of participation. *The Journal of Educational Research*, 91(3), 183-191.
- Mekinda, M., & Hirsch, B. (2010). Quality of work vs. non-work after-school activities for low-income, minority youth. Unpublished manuscript, Northwestern University.
- Moos, R. H. (1974). *The social climate scales: An overview*. Palo Alto, CA: Consulting Psychologists Press.
- Mortimer, J. (2003). *Working and growing up in America*. Cambridge, MA: Harvard University Press.
- Mortimer, J., Pimentel, E., Rye, S., Nash, K., & Lee, C. (1996). Part-time work and occupational value formation in adolescence. *Social Forces*, 74, 1305-18.
- Murnane, R., & Levy, F. (1996). *Teaching the new basic skills: principles for educating children to thrive in a changing economy*. New York: Free Press.
- National Research Council and Institute of Medicine (NRC/IOM) Committee on Increasing High School Students' Engagement and Motivation to Learn. (2003). *Engaging schools: Fostering high school students' motivation to learn*. Washington, DC: National Academy Press.
- National Research Council and Institute of Medicine (NRC/IOM) Committee on Community-Level Programs for Youth. (2002). *Community programs to promote youth development* (J. Eccles & J. Gootman, Eds.). Washington, DC: National Academy Press.
- National Research Council (NRC) Committee on Youth Employment Programs. (1985). *Youth employment and training programs: The YDPA years* (C. Betsey, R. G. Hollister Jr., & M. Papageorgiou, Eds.). Washington, DC: National Academy Press.
- Newman, K. S. (1999). *No shame in my game: The working poor in the inner city*. New York: Knopf and the Russell Sage Foundation.
- Noam, G., Biancarosa, G., & Dechausay, N. (2003). *Afterschool education: Approaches to an emerging field*. Cambridge, MA: Harvard Education Press.
- Owens, A., & Schneider, B. (2005). Self-regulation and the transition to adulthood. *Academic Exchange Quarterly*, 9, 62-67.

- Pedersen, S., & Seidman, E. (2004). Contexts and correlates of out-of-school activity participation among low-income urban adolescents. In J. L. Mahoney, R. W. Larson & J. S. Eccles (Eds.), *Organized activities as contexts of development: Extracurricular activities, after-school and community programs* (pp. 85-109). Mahwah, NJ: Lawrence Erlbaum Associates.
- Philliber, S., Kaye, J., Herrling, S., & West, E. (2002). Preventing pregnancy and improving health care access among teenagers: An evaluation of the Children's Aid Society – Carrera model. *Perspectives on Sexual and Reproductive Health, 34*, 244-251.
- Pittman, K., Irby, M., & Ferber, T. (2000). *Unfinished business: Further reflections on a decade of promoting youth development*. Philadelphia: Public/Private Ventures.
- Quinn, J. (1999). Where need meets opportunity: Youth development programs for early teens. *The Future of Children, 9*, 96-116.
- Quiroz, P., Gonzalez, N., & Frank, K. (1996). Carving a niche in the high school social structure: Formal and informal constraints on participation in the extra curriculum. *Research in Sociology of Education and Socialization, 11*, 93-120.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage Publications.
- Rhodes, J. (2002). *Stand by me: The risks and rewards of mentoring today's youth*. Cambridge, MA: Harvard University Press.
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.
- Rogoff, B. (2003). *The cultural nature of human development*. Oxford: Oxford University Press.
- Rosenbaum, J. (2001). *Beyond college for all: Career paths for the forgotten half*. New York: Russell Sage Foundation.
- Ruhm, C. (1997). Is high school employment consumption or investment? *Journal of Labor Economics, 15*, 735-776.
- Rumberger, R., & Larson, K. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education, 107*, 1-35.

- Sarason, S. (1972). *The creation of settings and the future societies*. Cambridge, MA: Brookline Books.
- Sarason, S. (1982). *The culture of the school and the problem of change*. Boston: Allyn & Bacon.
- Schank, R. C. (1995). *Dynamic memory revisited*. New York: Cambridge University Press.
- Seidman, E. (1988). Back to the future, community psychology: Unfolding a theory of social intervention. *American Journal of Community Psychology, 16*, 3-24.
- Shelton, S. H. (1990). Developing the construct of general self-efficacy. *Psychological Reports, 66*, 987-994.
- Sherer, M., Maddux, J., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports, 51*, 663-671.
- Simmons, R., & Blyth, D. (1987). *Moving into adolescence: The impact of pubertal change and school context*. New York: Aldine de Gruyter.
- Smith, C. & Hoffman, C. (2005). High/Scope Youth PQA technical report: Full findings from the Youth PQA Validation Study. Ypsilanti, MI: High/Scope Educational Research Foundation.
- Snyder, H., & Sickmund, M. (1999). *Juvenile offenders and victims: 1999 National Report*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Staff, J., Messersmith, E. E., & Schulenberg, J. E. (2009). Adolescents and the world of work. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., Vol. 2, pp. 270-313). Hoboken, NJ: Wiley.
- Tienda, M., & Ahituv, A. (1996). Ethnic differences in school departure: Does youth employment promote or undermine educational attainment? In G. Mangum & S. Mangum (Eds.), *Of heart and mind: Social policy essays in honor of Sar Levitan* (pp. 93-110). Kalamazoo, MI: Upjohn Institute.
- Undry, J. R. (1998). The National Longitudinal Study of Adolescent Health (ADD HEALTH), waves I & II, 1994-1996: A user's guide. Los Altos, CA: Sociometrics Corporation.
- Voelkl, K. E. (1996). Measuring students' identification with school. *Educational and Psychological Measurement, 56*, 760-770.

- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wood, R., & Bandura, A. (1989). Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. *Journal of Personal and Social Psychology*, 56, 407-415.
- Zief, S., Lauver, S., & Maynard, R. (2006). Impacts of after-school programs on student outcomes. *Campbell Systematic Reviews*, 3.
- Zimmerman, B. (1995 a). Self-regulation involves more than metacognition: A social cognitive perspective. *Educational Psychologist*, 30, 217-221.
- Zimmerman, B. (1995 b). Self-efficacy and educational development. In A. Bandura, (Ed.), *Self-efficacy in changing societies* (pp. 202-231). Cambridge, UK: Cambridge University Press.
- Zimmerman, B. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review*, 2, 173-201.

TABLES

TABLE 2.1
Data Collected for the Evaluation

	Measure	Source	Time of Assessment
Background/Control Variables	<ul style="list-style-type: none"> • Age, Gender, Ethnicity • SES: mother’s educational attainment, free/reduced lunch qualification • Extracurricular Activities 	<ul style="list-style-type: none"> • All background variables were collected using participant self-report computer survey 	<ul style="list-style-type: none"> • All variables were collected at the pretest survey
Implementation Variables	<ul style="list-style-type: none"> • Qualitative Observations • High Scope • Social Climate^a • Design Features^b 	<ul style="list-style-type: none"> • Weekly ASM program visits • Research staff ratings • Participant ratings of apprenticeship (treatment group) or extracurricular activities (control group) 	<ul style="list-style-type: none"> • Treatment duration • End of treatment • Posttest
Outcome Variables (by Domain)			
Positive Youth Development	<ul style="list-style-type: none"> • Self-efficacy • Self-regulation • Occupational Values • Relationships with Adults • Interpersonal Relationships 	<ul style="list-style-type: none"> • All positive youth development variables were collected using participant self-report computer survey 	<ul style="list-style-type: none"> • Pre & posttest • Pre & posttest • Pre & posttest • Pre & posttest • Posttest only
Marketable Job Skills	<ul style="list-style-type: none"> • Employability for generic, permanent, entry-level position 	<ul style="list-style-type: none"> • Mock-job interview, ratings completed by experienced HR professionals 	<ul style="list-style-type: none"> • Posttest
Academic	<ul style="list-style-type: none"> • Attitudes about School^b • Academic Performance • School Attendance 	<ul style="list-style-type: none"> • Participant self-report computer survey • Student transcripts & archival school records 	<ul style="list-style-type: none"> • Posttest • Academic year of study participation
Problem Behavior	<ul style="list-style-type: none"> • 10-item questionnaire 	<ul style="list-style-type: none"> • Participant self-report computer survey 	<ul style="list-style-type: none"> • Pre & posttest

^a Administered to Controls in Years 2 and 3 only. Administered to Treatment all three years.

^b Administered in Years 2 and 3 only.

TABLE 2.3
Treatment Attrition

Assigned to Treatment	No Show	Dropped during Fall	Dropped before Spring	Dropped during Spring	Total Dropout ^a
313	44	30	41	32	147
(% of total assigned to treatment)	(14.1%)	(9.6%)	(13.1%)	(10.2%)	(47.0%)

^a An additional 15 students (4.8% of the total assigned to treatment) participated in the apprenticeship through the end of the spring semester, but they did not meet the ASM standard attendance rate (73% or more). These students were not technically dropouts, but they were excluded from all treatment-on-the-treated analyses.

TABLE 2.4
Study Participation Rates

	Total	Assigned to Treatment	Control
Total N	547	313	234
Completed Posttest Survey	471	278	193
(% of total group assignment)	86.1%	88.8%	82.5%
School records	522	298	224
(% of total group assignment)	95.4%	95.2%	95.7%
Final N in Study	535	304	231

TABLE 2.5
Demographics: Entire Sample

	Total	Assigned to Treatment	Assigned to Control	F-test / Chi-Square	p-value
N	535	304	231	--	--
Mean age in years at pretest	15.87 (1.08)	15.85 (1.05)	15.90 (1.11)	.26	0.613
Grade:					
Freshman	36.4%	35.9%	37.2%		
Sophomore	29.5%	30.9%	27.7%	1.11	0.774
Junior	23.6%	23.7%	23.4%		
Senior	10.5%	9.5%	11.7%		
Female	59.4%	57.9%	61.5%	0.70	0.404
African American	76.6%	76.6%	76.6%	<0.01	0.995
Latino/a	22.5%	22.4%	22.8%	0.01	0.914
Reported prior ASM experience	20.5%	19.7%	21.8%	0.34	0.562
Participated in extracurricular activity in prior year ^a	82.4%	81.6%	83.5%	0.35	0.553
Reported prior job at pretest	74.8%	76.0%	73.2%	0.56	0.456
SES Components:					
Mother's Education: H.S. Grad	80.6%	78.0%	84.2%	3.11	0.078
Free/Reduced Price Lunch	92.2%	92.1%	92.5%	0.03	0.857
Reading & Math Composite	4.88	4.83	4.94		
Mean Stanine of 2 Prior Tests	(1.32)	(1.33)	(1.32)	0.90	0.342
Reading Standardized Tests	4.93	4.91	4.97		
Mean Stanine of 2 Prior Tests	(1.38)	(1.42)	(1.33)	0.25	0.621
Math Standardized Tests	4.83	4.76	4.92		
Mean Stanine of 2 Prior Tests	(1.48)	(1.45)	(1.52)	1.53	0.217

^a Extracurricular participation control entered in model was measured by amount of time spent in the activity. Groups did not significantly differ on this measure.

TABLE 2.6
Demographics: Treatment-on-the-Treated

	Total	Treatment^b	Control	F-test / Chi-Square	p-value
N	359	151	207	-	-
Mean age in years at pretest	15.85 (1.12)	15.78 (1.11)	15.90 (1.13)	1.00	0.317
Grade:					
Freshman	38.4%	39.5%	37.7%		
Sophomore	27.9%	28.9%	27.1%	0.57	0.902
Junior	22.8%	21.7%	23.7%		
Senior	10.9%	9.9%	11.6%		
Female	58.8%	55.9%	60.9%	0.89	0.347
African American	79.1%	82.9%	76.3%	2.29	0.131
Latino/a	20.3%	17.1%	22.7%	1.66	0.198
Reported prior ASM experience	19.1%	16.8%	20.9%	0.94	0.333
Participated in extracurricular activity in prior year ^a	81.1%	79.6%	82.1%	0.36	0.547
Reported prior job at pretest	73.8%	74.3%	73.4%	0.04	0.846
SES Components:					
Mother's Education: H.S. Grad	80.5%	77.4%	82.8%	1.74	0.420
Free/Reduced Price Lunch	93.2%	94.7%	92.1%	0.94	0.333
Reading & Math Composite Mean Stanine of 2 Prior Tests	4.92 (1.31)	4.90 (1.32)	4.94 (1.32)	0.09	0.763
Reading Standardized Tests Mean Stanine of 2 Prior Tests	4.95 (1.36)	4.93 (1.40)	4.98 (1.33)	0.07	0.793
Math Standardized Tests Mean Stanine of 2 Prior Tests	4.90 (1.48)	4.87 (1.47)	4.91 (1.51)	0.09	0.770

^a Extracurricular participation control entered in model was measured by amount of time spent in the activity. Groups did not significantly differ on this measure.

^b Treatment group defined by the ASM 73% attendance rule.

TABLE 2.7
Treatment-on-the-Treated/Dropout Pretest Demographics

	Treatment^a	Assigned to Treatment Dropout	p-value
N	151	153	
Demographics			
Mean age in years at pretest	15.78 (1.11)	15.92 (0.99)	0.238
Female	56.3%	59.5%	0.574
African American	82.8%	70.6%	0.012
Latino/a	17.2%	27.5%	0.032
Reported prior ASM experience	16.2%	23.1%	0.135
Participated in extracurricular activity in prior year	79.5%	83.7%	0.346
Reported prior job at pretest	74.8%	77.1%	0.640
SES Components:			
Mother's Education: H.S. Grad	77.2%	78.7%	0.768
Free/Reduced Price Lunch	94.7%	89.5%	0.095
Reading & Math Composite Mean Stanine of 2 Prior Tests	4.90 (1.29)	4.76 (1.36)	0.366
Pretest Measures			
Self-Efficacy	4.31 (0.50)	4.15 (0.54)	0.011
Self-Regulation	4.05 (0.59)	3.94 (0.58)	0.101
Adult Relationships	3.92 (0.54)	4.00 (0.56)	0.514
Occupational Values	4.16 (0.54)	4.08 (0.64)	0.234
Problem Behaviors	1.13 (0.25)	1.12 (0.20)	0.812

^aTreatment group defined by the ASM 73% attendance rule.

TABLE 2.8
Control Group Extracurricular Activities

Most time-intensive activity (N = 112)	
Sports team	33.0%
Performance, music, or art program	25.0%
Service group or club	10.7%
Academic or leadership club	10.7%
Community program	8.0%
Paid job	12.5%

TABLE 3.1
Implementation Measure: High/Scope

Subscales		High/Scope Validation Sample ^a N = 118 Programs	Participating Apprenticeships N = 13 Programs
Supportive Environment	Mean (sd)	3.77 (0.83)	3.80 (0.62)
	Min	4.68	2.70
	Max	5.00	4.70
Interaction	Mean (sd)	3.03 (0.90)	3.45 (0.65)
	Min	1.00	2.04
	Max	4.83	4.34
Engagement	Mean (sd)	2.68 (1.11)	3.46 (0.92)
	Min	1.00	1.94
	Max	5.00	4.83

^a Smith & Hoffman (2005).

TABLE 3.2
Implementation Measure: Design Features

	Treatment N = 114 Mean (sd)	Control N = 112 Mean (sd)	Effect Size (g)	F-Test	p-value
Design Features - Total	4.39 (0.62)	3.80 (0.70)	0.89	42.77	<.001
-Adult taught skill	4.59 (0.69)	3.95 (1.20)	0.66	23.93	<.001
-Deadline	4.20 (1.05)	3.64 (1.23)	0.49	13.02	<.001
-Adult told you how to improve	4.39 (0.95)	3.72 (1.36)	0.57	17.77	<.001
-Related to job/career	4.35 (0.88)	3.25 (1.47)	0.91	45.23	<.001
-Worked in groups	4.41 (1.05)	4.33 (1.15)	0.07	0.24	0.625
-Youth choice in activities	4.37 (0.82)	3.75 (1.09)	0.64	22.99	<.001
-Project viewed by others	4.39 (0.97)	3.99 (1.18)	0.37	7.39	0.007

Sample sizes for these analyses reflect that this measure was given to both groups during Years 2 and 3 only.

TABLE 3.3
Implementation Measure: Social Climate

	Treatment N = 140 Mean (sd)	Control N = 112 Mean (sd)	Effect Size (g)	F-Test	p-value
Social Climate - Total Scale	3.94 (0.55)	3.87 (0.54)	0.13	0.81	0.369
-Cohesion	3.96 (0.78)	3.92 (0.77)	0.05	0.16	0.692
-Task Orientation	4.01 (0.64)	3.94 (0.67)	0.11	0.46	0.500
-Leader Support	4.19 (0.68)	4.14 (0.71)	0.07	0.31	0.581

Sample sizes for these analyses reflect that this measure was given to both groups during Years 2 and 3 only.

Table 3.4
Implementation and Outcome Measure Correlations

	Design Features Total	Social Climate Total
Positive Youth Development		
Self-Efficacy	0.21(**)	0.61(**)
Self-Regulation	0.25(**)	0.57(**)
Adult Relationships	0.17(**)	0.44(**)
Occupational Values	0.08	0.29(**)
Interpersonal Learning	0.26(**)	0.50(**)
Hiring		
Likelihood of hiring for permanent job	0.06	0.13(*)
Composite index of 13 interview questions	0.09	0.16(**)
Composite index of applicant qualities	0.09	0.13(**)
School Identification		
School Identification Total Scale	0.25(**)	0.53(**)
School Identification - Extrinsic Value	0.26(**)	0.46(**)
School Identification - Belonging	0.19(**)	0.47(**)
Academic Performance		
Weighted GPA - Whole Year	0.07	0.12(*)
# Failed Courses - Whole Year	-0.07	-0.08
# Days Absent - Whole Year	-0.03	0.02
Problem Behavior		
Total Scale Mean	0.02	-0.09

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

^a Sample sizes range from 323-334 (combined Treatment and Control). Design features data collected in Years 2 and 3 only.

^b Sample sizes range from 333-371 (combined Treatment and Control). Social climate data collected from Controls in Years 2 and 3 only.

TABLE 4.1
Positive Youth Development Outcomes
Intent-to-Treat

Measure	Treatment			Control			Effect Size ^a (g)	Treatment T-Ratio	p-value
	N	Pretest Mean (sd)	Posttest Mean (sd)	N	Pretest Mean (sd)	Posttest Mean (sd)			
Self-Efficacy	256	4.23 (.52)	4.09 (.58)	175	4.23 (.51)	4.05 (.62)	0.07	0.93	0.353
Self-Regulation	255	3.99 (.59)	3.88 (.62)	173	4.01 (.54)	3.79 (.62)	0.18	2.14	0.033
Adult Relationships	257	3.95 (.56)	3.73 (.56)	174	3.88 (.54)	3.71 (.59)	-0.09	-0.38	0.705
Occupational Values	261	4.13 (.57)	4.06 (.60)	180	4.14 (.52)	4.10 (.58)	-0.05	-0.37	0.710
Interpersonal Learning	239	--	4.24 (.64)	150	--	4.22 (.72)	0.03	0.83	0.406

^aEffect sizes are based on unadjusted means in this and all subsequent tables.

TABLE 4.2
Marketable Job Skills Outcomes
Intent-to-Treat

	Treatment N = 262 Mean (sd)	Control N = 182 Mean (sd)	Effect Size (g)	Treatment T-ratio	p-value
Hired	50.8%	53.3%	–	–	–
Likelihood of hiring for permanent job	3.33 (1.31)	3.42 (1.24)	-0.07	-0.21	0.831
Likelihood of hiring for summer job	3.67 (1.23)	3.79 (1.15)	-0.10	-0.43	0.671
Composite index of 13 interview questions	3.92 (.58)	3.93 (.57)	-0.02	0.17	0.863
Composite index of applicant qualities	3.86 (.76)	3.89 (.72)	-0.04	-0.13	0.901
Q1A - Why applied	3.63 (1.09)	3.51 (1.11)	0.11	1.56	0.120
Q2A - What experiences do you have	3.49 (1.32)	3.54 (1.28)	-0.04	0.10	0.922
Q3A - Recent goal	3.60 (1.25)	3.70 (1.28)	-0.08	-0.14	0.887
Q4A - Working with others	4.01 (1.10)	4.15 (1.05)	-0.13	-0.98	0.327
Q5A - Complete a project by deadline	4.07 (1.12)	4.09 (1.10)	-0.02	0.02	0.987
Q6A - Situation dealing with angry person	3.55 (1.39)	3.69 (1.31)	-0.10	-1.22	0.223
Q7A - Disagree with supervisor	4.48 (.93)	4.45 (.88)	0.03	0.11	0.916
Q8A - Stay late to cover shifts	4.53 (.87)	4.43 (.92)	0.11	0.99	0.322
Q9A - Special event conflict	4.38 (1.05)	4.45 (.97)	-0.07	-1.01	0.315
Q10A - Trouble with photocopier	4.54 (.92)	4.50 (.90)	0.04	0.56	0.576
Q11A - How job contributes to future	3.77 (1.24)	3.73 (1.32)	0.03	1.11	0.266
Q12A - Questions about job	3.31 (1.71)	3.30 (1.63)	0.01	0.11	0.915
Q13A - What else about you	3.67 (.87)	3.60 (.96)	0.08	0.46	0.645
Q1B - Initial impression	3.47 (1.09)	3.51 (1.03)	-0.04	-0.37	0.710
Q2B - Appropriate dress and appearance	3.84 (.97)	3.75 (1.00)	0.09	0.78	0.437
Q3B - Eye contact	4.11 (.96)	4.03 (1.04)	0.08	0.63	0.527
Q4B - Positive attitude	3.91 (1.07)	3.96 (1.04)	-0.05	-0.41	0.682
Q5B - Body language	4.16 (.95)	4.19 (.95)	-0.03	-0.23	0.819
Q6B - Confidence	3.70 (1.14)	3.82 (1.10)	-0.11	-0.98	0.329
Q7B - Paid attention	4.35 (.85)	4.42 (.77)	-0.09	-0.93	0.355
Q8B – Communication	3.67 (1.17)	3.66 (1.12)	0.01	0.64	0.524
Q9B – Maturity	3.89 (1.21)	4.04 (1.12)	-0.13	-0.93	0.353
Q10B - Application completion	3.72 (1.19)	3.65 (1.18)	0.06	1.22	0.226
Q11B - Prior experience	3.09 (1.20)	3.16 (1.27)	-0.06	-0.35	0.729
Q12B - Response to feedback	4.37 (.73)	4.43 (.73)	-0.08	-0.43	0.669

TABLE 4.3
Academic Performance Outcomes
Intent-to-Treat

	Treatment N = 270 Mean (sd)	Control N = 192 Mean (sd)	Effect Size (g)	Treatment T-Ratio	p-value
Weighted GPA - Whole Year	2.27 (.95)	2.32 (.98)	-0.05	-0.73	0.463
Weighted GPA - Fall	2.32 (.97)	2.38 (.99)	-0.06	-0.93	0.353
Weighted GPA - Spring	2.21 (1.03)	2.25 (1.05)	-0.04	-0.39	0.694
# of failed courses - Whole Year	1.08 (1.83)	1.10 (2.01)	-0.01	0.32	0.751
# of failed courses - Fall	.47 (.95)	.46 (.94)	0.01	0.73	0.468
# of failed courses - Spring	.61 (1.09)	.64 (1.21)	-0.03	-0.24	0.813

TABLE 4.4
School Attendance with Prior Year Attendance Control
Intent-to-Treat

Absences	Treatment N = 82		Control N = 111		Effect Size (g)	Treatment T-Ratio	p-value
	Prior Year Mean (sd)	Current Year Mean (sd)	Prior Year Mean (sd)	Current Year Mean (sd)			
# of days - Whole Year	28.13 (19.45)	32.07 (19.84)	26.68 (21.07)	32.58 (21.08)	-0.10	-0.72	0.472
# of days - Fall	10.96 (8.62)	12.26 (9.46)	10.57 (10.45)	12.53 (9.49)	-0.07	-0.41	0.682
# of days - Spring	17.23 (12.35)	19.81 (12.13)	16.10 (11.90)	20.05 (13.20)	-0.11	-0.84	0.404

TABLE 4.5
School Attendance Entire Sample
Intent-to-Treat

Absences	Treatment N = 269 Mean (sd)	Control N = 192 Mean (sd)	Effect Size (g)	Treatment T-Ratio	p-value
# of days - Whole Year	29.37 (20.11)	29.75 (22.52)	-0.02	-0.11	0.916
# of days - Fall	11.23 (8.88)	11.58 (9.95)	-0.04	-0.51	0.608
# of days - Spring	18.14 (12.68)	18.17 (13.90)	-0.002	0.05	0.964

TABLE 4.6
Attitudes about School
Intent-to-Treat

	Treatment N = 220 Mean (sd)	Control N = 145 Mean (sd)	Effect Size (g)	Model 1		Model 2	
				Treatment T-Ratio	p-value	Treatment T-Ratio	p-value
School Identification	4.05 (.46)	3.94 (.50)	0.23	2.22	0.027	1.89	0.059
-Extrinsic Value	4.21 (.51)	4.10 (.55)	0.21	1.97	0.050	1.81	0.071
-Belonging	3.91 (.56)	3.81 (.59)	0.18	1.82	0.069	1.44	0.150

Sample sizes for these analyses reflect that this measure was given during Years 2 and 3 only.

TABLE 4.7
Problem Behavior Outcomes
Intent-to-Treat

Item	Treatment N = 259		Control N = 178		Effect Size (g)	Treatment T-Ratio	p-value
	Pretest Mean (sd)	Posttest Mean (sd)	Pretest Mean (sd)	Posttest Mean (sd)			
Total Scale Mean	1.13 (.22)	1.20 (.30)	1.16 (.28)	1.28 (.53)	-0.12	-1.95	0.051
Drink Alcohol	1.17 (.51)	1.29 (.66)	1.18 (.57)	1.36 (.79)	-0.08	-1.30	0.194
Use Drugs	1.03 (.20)	1.17 (.63)	1.09 (.34)	1.31 (.93)	-0.10	-1.22	0.224
Risky Intercourse	1.22 (.75)	1.36 (.96)	1.29 (.81)	1.40 (.93)	0.03	-0.66	0.510
Physical Fights	1.38 (.73)	1.51 (.88)	1.38 (.77)	1.53 (.94)	-0.02	-0.70	0.486
Steal <\$50	1.05 (.27)	1.11 (.42)	1.08 (.33)	1.20 (.66)	-0.11	-1.51	0.132
School Suspension	1.27 (.63)	1.36 (.72)	1.29 (.64)	1.40 (.75)	-0.03	-1.05	0.297
Sell Drugs	1.03 (.28)	1.03 (.21)	1.07 (.49)	1.16 (.70)	-0.19	-2.61	0.010
Steal > \$50	1.01 (.11)	1.03 (.16)	1.04 (.22)	1.09 (.55)	-0.08	-1.83	0.067
Carry a Weapon	1.05 (.32)	1.09 (.49)	1.05 (.27)	1.16 (.65)	-0.12	-1.36	0.174
Gang Activity	1.05 (.34)	1.07 (.39)	1.07 (.42)	1.18 (.78)	-0.15	-2.01	0.044

TABLE 5.1
Positive Youth Development Outcomes
Treatment-on-the-Treated

Measure	Treatment			Control			Effect Size (g)	Treatment T-Ratio	p-value
	N	Pretest Mean (sd)	Posttest Mean (sd)	N	Pretest Mean (sd)	Posttest Mean (sd)			
Self-Efficacy	132	4.30 (.51)	4.14 (.57)	163	4.24 (.51)	4.05 (.63)	0.05	0.82	0.411
Self-Regulation	131	4.05 (.60)	3.95 (.65)	161	4.01 (.55)	3.79 (.64)	0.19	2.15	0.032
Adult Relationships	133	3.92 (.54)	3.75 (.58)	162	3.88 (.54)	3.71 (.59)	0.00	0.15	0.878
Occupational Values	135	4.15 (.53)	4.00 (.66)	167	4.15 (.52)	4.10 (.59)	-0.16	-1.22	0.224
Interpersonal Learning	134	–	4.25 (.65)	138	–	4.20 (.73)	0.07	1.02	0.308

TABLE 5.2
Marketable Job Skills Outcomes
Treatment-on-the-Treated

	Treatment N = 135 Mean (sd)	Control N = 169 Mean (sd)	Effect Size (g)	Treatment T-ratio	p-value
% Hired	56.3%	54.4%	–	–	–
Likelihood of hiring for permanent job	3.47 (1.26)	3.44 (1.26)	0.02	1.08	0.282
Likelihood of hiring for summer job	3.83 (1.16)	3.79 (1.15)	0.04	1.23	0.220
Composite index of 13 interview questions	4.01 (.52)	3.95 (.56)	0.11	1.97	0.049
Composite index of applicant qualities	3.97 (.71)	3.88 (.73)	0.13	1.76	0.079
Q1A - Why applied	3.68 (1.03)	3.52 (1.12)	0.15	1.79	0.075
Q2A - What experiences do you have	3.61 (1.32)	3.53 (1.27)	0.06	1.44	0.152
Q3A - Recent goal	3.65 (1.23)	3.75 (1.27)	-0.08	0.48	0.631
Q4A - Working with others	4.13 (1.03)	4.16 (1.06)	-0.03	0.07	0.947
Q5A - Complete a project by deadline	4.14 (1.08)	4.08 (1.12)	0.05	0.84	0.403
Q6A - Situation dealing with angry person	3.59 (1.40)	3.73 (1.28)	-0.11	-1.22	0.825
Q7A - Disagree with supervisor	4.63 (.70)	4.44 (.89)	0.23	2.20	0.029
Q8A - Stay late to cover shifts	4.53 (.81)	4.44 (.89)	0.11	1.14	0.256
Q9A - Special event conflict	4.48 (.95)	4.44 (.95)	0.04	0.32	0.753
Q10A - Trouble with photocopier	4.64 (.83)	4.49 (.89)	0.17	1.52	0.129
Q11A - How job contributes to future	3.84 (1.25)	3.79 (1.27)	0.04	0.81	0.418
Q12A - Questions about job	3.47 (1.67)	3.28 (1.63)	0.12	1.37	0.172
Q13A - What else about you	3.71 (.86)	3.63 (.93)	0.09	1.12	0.266
Q1B - Initial impression	3.57 (1.06)	3.52 (1.04)	0.05	0.82	0.414
Q2B - Appropriate dress and appearance	3.98 (.97)	3.73 (1.00)	0.25	1.73	0.085
Q3B - Eye contact	4.21 (.97)	4.02 (1.05)	0.19	1.57	0.118
Q4B - Positive attitude	4.01 (.97)	3.97 (1.04)	0.04	1.05	0.295
Q5B - Body language	4.30 (.86)	4.19 (.96)	0.12	1.13	0.259
Q6B - Confidence	3.83 (1.08)	3.83 (1.10)	<0.01	0.45	0.650
Q7B - Paid attention	4.44 (.80)	4.41 (.78)	0.04	0.86	0.393
Q8B – Communication	3.79 (1.07)	3.65 (1.12)	0.13	1.89	0.059
Q9B – Maturity	4.01 (1.10)	4.06 (1.13)	-0.05	0.42	0.677
Q10B - Application completion	3.84 (1.22)	3.63 (1.18)	0.18	2.44	0.015
Q11B - Prior experience	3.22 (1.14)	3.15 (1.29)	0.06	1.28	0.201
Q12B - Response to feedback	4.43 (.72)	4.42 (.74)	0.01	0.84	0.402

TABLE 5.3
Academic Performance Outcomes
Treatment-on-the-Treated

	Treatment N = 134 Mean (sd)	Control N = 177 Mean (sd)	Effect Size (g)	Treatment T-Ratio	p-value
Weighted GPA - Whole Year	2.41 (.98)	2.31 (.99)	0.10	1.04	0.298
Weighted GPA - Fall	2.47 (1.01)	2.38 (1.00)	0.09	0.77	0.445
Weighted GPA - Spring	2.36 (1.04)	2.25 (1.05)	0.11	1.29	0.199
# of failed courses - Whole Year	.83 (1.69)	1.01 (2.01)	-0.10	-1.01	0.313
# of failed courses – Fall	.34 (.82)	.46 (.93)	-0.14	-0.58	0.565
# of failed courses – Spring	.49 (1.01)	.64 (1.22)	-0.13	-1.24	0.217

TABLE 5.4
School Attendance with Prior Year Attendance Control
Treatment-on-the-Treated

Absences	Treatment N = 82		Control N = 111		Effect Size (g)	Treatment T-Ratio	p-value
	Prior Year Mean (sd)	Current Year Mean (sd)	Prior Year Mean (sd)	Current Year Mean (sd)			
# of days - Whole Year	24.24 (16.11)	28.33 (16.45)	27.58 (21.51)	32.42 (21.51)	-0.04	-0.81	0.421
# of days - Fall	9.80 (7.13)	10.27 (7.06)	11.03 (10.73)	12.16 (9.58)	-0.08	-0.46	0.647
# of days - Spring	14.57 (10.18)	18.06 (10.99)	16.55 (12.06)	20.27 (13.43)	-0.02	-0.91	0.365

TABLE 5.5
School Attendance Entire Sample
Treatment-on-the-Treated

Absences	Treatment N = 133 Mean (sd)	Control N = 177 Mean (sd)	Effect Size (g)	Treatment T-Ratio	p-value
# of days - Whole Year	25.68 (17.53)	29.72 (23.01)	-0.19	-1.68	0.094
# of days - Fall	9.42 (7.26)	11.37 (10.12)	-0.22	-2.02	0.044
# of days - Spring	16.26 (11.63)	18.35 (14.17)	-0.16	-1.43	0.154

TABLE 5.6
Attitudes about School
Treatment-on-the-Treated

	Treatment N = 106 Mean (sd)	Control N = 136 Mean (sd)	Effect Size (<i>g</i>)	Model 1		Model 2	
				Treatment T-Ratio	p-value	Treatment T-Ratio	p-value
School Identification	4.08 (.47)	3.94 (.51)	0.28	2.25	0.025	2.30	0.023
-Extrinsic Value	4.27 (.54)	4.09 (.55)	0.33	2.65	0.009	2.74	0.007
-Belonging	3.91 (.55)	3.81 (.60)	0.17	1.35	0.180	1.34	0.182

Sample sizes for these analyses reflect that this measure was given during Years 2 and 3 only.

TABLE 5.7
Problem Behavior Outcomes
Treatment-on-the-Treated

Item	Treatment N = 134		Control N = 166		Effect Size (g)	Treatment T-Ratio	p-value
	Pretest Mean (sd)	Posttest Mean (sd)	Pretest Mean (sd)	Posttest Mean (sd)			
Total Scale Mean	1.12 (.22)	1.18 (.30)	1.16 (.28)	1.28 (.54)	-0.13	-1.56	0.122
Drink Alcohol	1.21 (.49)	1.27 (.63)	1.18 (.59)	1.36 (.77)	-0.17	-1.84	0.066
Use Drugs	1.03 (.21)	1.16 (.62)	1.09 (.35)	1.30 (.91)	-0.10	-0.68	0.498
Risky Intercourse	1.17 (.61)	1.32 (.91)	1.30 (.84)	1.41 (.95)	0.04	-0.40	0.689
Physical Fights	1.36 (.63)	1.40 (.76)	1.37 (.77)	1.54 (.96)	-0.15	-1.68	0.094
Steal <\$50	1.05 (.31)	1.12 (.41)	1.09 (.35)	1.19 (.67)	-0.05	-0.56	0.574
School Suspension	1.25 (.54)	1.26 (.57)	1.30 (.66)	1.40 (.75)	-0.13	-2.00	0.046
Sell Drugs	1.05 (.37)	1.04 (.23)	1.08 (.51)	1.17 (.73)	-0.18	-1.96	0.051
Steal > \$50	1.00 (.00)	1.03 (.17)	1.04 (.23)	1.10 (.56)	-0.07	-1.36	0.170
Carry a Weapon	1.08 (.42)	1.11 (.55)	1.05 (.28)	1.17 (.67)	-0.15	-0.82	0.411
Gang Activity	1.04 (.29)	1.08 (.46)	1.07 (.43)	1.17 (.77)	-0.09	-1.29	0.200

TABLE 6.1
Average Effect Size by Outcome Domain

Domain	ASM ITT ^a	ASM TOT ^b	Durlak et al. (2010) meta-analysis		
			All 68 programs	SAFE	Non- SAFE
Positive Youth Development (Self-Perceptions)	.03	.03	.34	.37	.13
Marketable Job Skills	-.02	.08	–	–	–
Academics	–	–	–	–	–
Grades	-.02	.12	.12	.22	.05
Attendance	.09	.05	.10	.14	.07
School Identification (Bonding)	.20	.25	.14	.25	.03
Problem Behavior	.09	.11	.19	.30	.08

^a Average effect sizes calculated using the intent-to-treat analysis.

^b Average effect sizes calculated using the treatment-on-the-treated analysis.