Document Title:	**Final Technical Report: An Evaluation of Advocacy-based Mentoring as a Treatment Intervention for Chronic Delinquency**
Author(s):	**Michael J. Karcher, Ed.D., Ph.D., David A. Johnson, Ph.D.**
Document Number:	**250454**
Date Received:	**December 2016**
Award Number:	**2011-JU-FX-0001**

Final Technical Report

AN EVALUATION OF ADVOCACY-BASED MENTORING AS A TREATMENT INTERVENTION FOR CHRONIC DELINQUENCY

The U.S. Department of Justice, Office of Justice Programs
Office of Juvenile Justice and Delinquency Prevention
OJJDP FY 2011 Research on Best Practices for Mentoring

Michael J. Karcher, Ed.D., Ph.D.
Professor and Principal Investigator
Department of Educational Psychology
University of Texas, San Antonio
210-458-2302

with assistance from
David A. Johnson, Ph.D., Assistant Professor
Communication Disorders and Counseling, School, and Educational Psychology
Bayh College of Education
Indiana State University

Text for the first study, section one, was initially drafted by Dr. Mitch Miller.

Disclaimer: This project was supported by Grant No. (2011-JU-FX-0001) awarded by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. Points of view or opinions in this document are those of the authors and do not necessarily represent the official position or polices of the U.S. Department of Justice.

Acknowledgements: Drs. Karcher and Johnson would like to acknowledge the great assistance provided by Martin J. D'Urso, Esq, Alan Kassirer, and Dr. Michael Marks at Youth Advocate Programs, Inc. and the Directors, Staff and Advocates at the Camden, NJ, Mobile, AL, Las Vegas, NV, Lebanon, PA, and Toledo, OH Youth Advocate Programs for their assistance in working with youth, parents and referring authorities to secure access to necessary data. Finally, we would like to thank Jennifer Tyson**,** Social Science Analyst, Innovation and Research Division at the Office of Juvenile Justice and Delinquency Prevention for her support and direction throughout the project.

Study Overview and Conclusions

The primary goal of the proposed research project was to provide estimates of the effectiveness of youth advocacy in general, and more specifically as delivered through the Youth Advocate Programs, Inc. (YAP). YAP is a national nonprofit organization active in twenty states that provides a treatment intervention for reducing serious and chronic delinquency for court-referred youth. This study examined processes and outcomes in the YAP program in four cities to inform juvenile justice policy and practice regarding the possible benefits of advocacy-based interventions for this population.

This grant focused on evaluating four Youth Advocate Programs, Inc. (YAP) in separate regions of the country in order to increase variability in model delivery and youth participants. The research design was based on information collected about YAP services from prior research and focused both on identifying key mentoring and advocacy processes that may interrupt chronic delinquency and measuring proximal and distal outcomes related to crime and prosocial behavior of participation in youth mentoring with paid mentors who prioritized advocacy as one element of mentoring.

Project Goals to achieve these objectives were two-fold: (1) Estimate the degree to which intended program objectives were realized (i.e., program impact and effectiveness) through attempts to quantify the association between participation in the YAP program and changes in youth delinquency and on related outcomes using a rigorous quasi-experimental research; and (2) identify ways in which advocacy and specific types of mentoring interactions contribute to youth outcomes through program participation. The overall goal of this study was, therefore, to better understand the viability of advocacy as an intervention for youth at high risk for future criminal activity, to identify critical practices that may be relevant to YAP and other programs using individualized treatment approaches to reduce delinquency and recidivism through advocacy efforts, and to learn more about which interpersonal interactions and participant characteristics are most influential in successful advocacy efforts.

Two adaptations to the originally proposed methods and design of the project were necessitated by factors and events beyond the control of the report authors. A formative evaluation of each program's fidelity of implementation was omitted from the final report because the data on which initial findings were to be based were not available once unforeseeable changes in project staffing occurred. Second, initially the quasi-experimental method for detecting effects associated with program participation was propensity score analyses. Data collected by a consultant on the project, however, were unsuitable for propensity score analyses. Therefore, this report does not include results related to implementation fidelity or propensity score analyses of causal effects. In short, the data was insufficient to cross check findings or reach reliable conclusions. However, another quasi-experimental design was used to estimate program causal effects, thereby allowing Study 1 to address the question about program impact. Study 2 uses program activities and participant characteristics to try to explain the changes reported in Study 1.

Study 1: Outcome Evaluation

For the quasi-experimental outcome analyses, the *recurrent institutional cycle* (RIC) quasi-experimental design was employed (Campbell & Stanley, 1963)[1] instead of the planned propensity score analyses because we found we lacked sufficient data to properly match youth on key variables, those most characteristic of YAP participants. As one example, the state-level data collected to match youth in the court system with youth participating in the YAP program in Las Vegas did not include information on youth's involvement in sex trafficking. Yet this youth characteristic is disproportionately higher among court referrals to YAP in Las Vegas. With very limited demographic background information such as this reliable matching could not be achieved, because it would mask problematically large unobserved differences between YAP's court-referred youth and the comparison group to whom their outcomes would be compared. Lacking confidence that any attributions of program-related outcomes could be generated from this data, an alternative design, one Campbell and Stanley identified as a highly rigorous quasi-experimental design, was selected for addressing the most problematic internal validity threat in quasi-experimental causal effect analyses—selection.

Statistical analysis of program effects on recidivism, deinstitutionalization, and pro-social activity (e.g., school and employment status) were estimated using the *recurrent institutional cycle* (RIC) design as described first by Campbell and Stanley (1963). The RIC design minimizes selection threats by comparing program graduates (i.e., the treatment group) to a counterfactual untreated group (i.e., comparison group) reflecting the pre-treatment data of the youth collected prior to their participation in YAP. This method can provide the closest approximation of the counterfactual state of what a YAP participant at a given age would look like had that person actually not participated in YAP. Campbell and Stanley argue that in some situations the counterfactual group in the RIC design is better than the counterfactual group derived from random assignment.

What this design does not allow, however, is the analysis of differences between the treatment and the untreated comparison group on variables not collected prior to entering the program. This is because, where no pre-intervention assessments of phenomenon are available, the intervention group has no pre-data for this outcome variable. For example, recidivism was not a viable outcome variable because the data we received from the states did not include dates of adjudication needed to the estimate recidivism relative to a specific determination, such as a referral to YAP or other alternative program.

However, there was considerable other data available assessing a variety of crime and recidivism-related phenomena to undertake the *recurrent institutional cycle* (RIC) analyses. Unique to the RIC design, we relied on the data regularly collected by YAP to

[1] Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Dallas, TX: Houghton Mifflin Company.

assess the graduates' involvement in crime subsequent to graduation, as this variable was collected prior to treatment for all youth, then at discharge, and finally at 12 months post treatment for program completers.

Data collection for the 163 participating youth in the study ended June 15, 2014. Of these, 163 youth completed pre-surveys, 133 youth and Advocate pairs completed 2-month surveys, and 103 youth and Advocate pairs completed 4-month surveys. To address the issue of attrition, we used multiple imputation to estimate missing data on participant outcomes at the 2-month and 4-month time point as well as for absent follow up data for those who could not be located a year post discharge.

Conclusions about program related changes. Findings from Study 1 suggest that participation in the YAP program was related to several self-reported outcomes, including improvements in academic connectedness and declines in self-reported misconduct. Equally large and consistent were improvements on disposition (crime engagement), educational engagement, and pursuit of employment were present at discharge. Furthermore, several benefits of program participation found at discharge were maintained 12 months post-discharge.

The RIC research design used for Study 1 requires a wide range of tests to rule out rival hypotheses, and wading through these analyses can be tedious. Ultimately it is the comparisons between treated and untreated youth that are of greatest interest to most readers. The results of these analyses revealed several differences favoring those who had participated in the YAP program in addition to those reported in the prior paragraph. Cross-cohort as well as pooled sample tests of within and between-group differences were computed. Subsequently, separate whole-sample hierarchical analyses (Mplus two-level analysis) were conducted in order to pool all youth and make treated-untreated group comparisons, controlling for the nestedness of the data within persons and by accounting for missing data using FIML, of change related to program participation.

The main effect results suggested that participation in the YAP program for the youth in this sample was associated with increases on several measures of prosocial behavior. Pooling the sample provided just barely sufficient statistical power to detect the expectably small effect sizes typical of mentoring (d = -.20), so a p-value of .10 was considered indicative of a non-chance finding and .20 as a trend, just as was done in the original impact study conducted on the Big Brothers Big Sisters community based mentoring program in the 1990s. Given these cut off scores, we found that the treated youth had significantly higher connectedness to teachers and a trend toward higher connectedness to school and a self-in-the-future.

The most compelling evidence of change associated with participation in YAP was the reduction in misconduct (self-reported) after program completion. Whole sample tests were corroborated by smaller-sample comparisons between treated and untreated youth in the between cohort comparisons. For cross-cohort comparisons the typical effect size for mentoring (d = -.20) was used instead of statistical probabilities as benchmark for

assessing the presence and absence of meaningful change. In these comparisons, four of the five cohorts showed small (d = -.20) to moderate (d = -.45) declines in misconduct. Further corroborating evidence of program-related change is that the average pre-post change from entry to discharge on misconduct (d = -.25) and was comparable in size to the average the between-groups/cross-cohort differences (d = .20). Furthermore, these effect sizes reflect changes associated with program participation that are comparable to or larger than the effects of most mentoring programs.[2]

Relatively large, positive changes also were observed both at discharge and 12-months post-discharge on criminality, educational engagement, and employment status. From entry to 12-months post-discharge, the severity of YAP youths' disposition (status offense, misdemeanor, felony) was lower, their school participation (e.g., attendance) was higher, and they reported more effort to secure employment than untreated youth of the same age who had not yet begun YAP.

Ib. Study 2: Program Practices and Advocate Characteristics

Using structural equation modeling, data from three time points (entry, two and four months, which is the average time of discharge) were included in path analyses to test predictions about the role of specific Advocate and mentoring interactions derived from the TEAM framework (Karcher & Nakkula, 2010). These were primarily that play and serious problem-focused conversations would explain rates of misconduct reported at discharge.

Structural models were utilized so that the different types of interactions engaged in between the start of the match through the second month and then between two and four months could serve as separate, unique predictors and moderators of outcomes, adjusting for the contribution to each made by other variables in the model, specifically the potentially confounding background characteristics of youth. Additional variables were subsequently included in these models to factor in the direct and indirect contributions made by both youth and Advocate characteristics in explaining changes in misconduct.

The first part of Study 2 looked at changes that occurred among youth during their time in YAP on self-reported misconduct as misconduct, which was the variable on which the most compelling evidence of program-related changes was found in Study 1. Of course to make inferences regarding program practices from these analyses, the reports of misconduct by youth, both at entry and at discharge, must be sufficiently accurate, reliable and valid indicators of true misconduct. Fortunately, the corroboration of findings in Study 1 using the self-reported misconduct measure with findings using

[2] DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology*, *30*, 157-197.

other indicators of misconduct, combined with the stability of findings over time on the other indicators (between discharge and 12 month follow-up reports), provide convergent and predictive validity evidence to support the utility of the self-reported misconduct scale as the primary outcome measure in Study 2.

The first analysis simply assessed whether or not change in self-reported misconduct was evident across entry, the midpoint (two months), and discharge. This structural model provided evidence of change across each time point that corroborated the findings from Study 1. Following these analyses we then conducted a series of structural models to test TEAM framework-based hypotheses about the way in which activities might explain these changes.

Structural path model analyses focused primarily on the relative contributions of playful/recreational activities, casual conversations, and problem-oriented discussions engaged in by Advocates with their mentees during the first and second half of their time together. The baseline model focused on rates of activities during the first and second half of the relationship separately. This allowed us to estimate how differences in what happened in the first and second half of the relationship may have uniquely contributed to variability in rates of misconduct among participating youth at discharge beyond what would have been predicted from youths' starting levels of misconduct. Next, youth demographic characteristics were factored into these models.

We then incorporated into these baseline path models two Advocate characteristics to see if the role of activities on program outcomes was influenced by Advocates' background characteristics and to see if these specific Advocate characteristics contributed directly to program-related outcomes. These path models also included control variables to account for variability associated with youths' age and sex. Using two final path models we compared the main and indirect effects of mentor education and prior teaching experience on program-related changes in youths' misconduct.

Two sets of path models were conducted. A first was run using observed (mean) score assessments of activity frequency. We felt these would make the translation or use of the findings by practitioners who may subsequently choose to use the same activity logs in their work easier. A second set of analyses was conducted using confirmatory factor analysis derived activity frequency scores. Using factor scores helped account for error within observed score assessments and allowed for a clearer view of what specific interactions each activity factor score reflected most. These models used problem-focused discussion frequencies while the observed score models used frequencies of casual conversations as the complementary activity to play in our hypothesized model tests. This was done for both theoretical and statistical reasons explained in the results section.

Overview of Findings. The path model analyses presented in Study 2,

particularly the final observed score model that incorporated relevant youth characteristics (considered potential confounds) revealed that the benefits of program participation varied as a consequence of the frequency of activities at specific time points in the relationship and as a function of at least these two Advocate characteristics (prior teaching experience and educational attainment), both of which contributed to program outcomes directly and indirectly.

First, the models revealed that higher rates of play and lower rates of both casual conversations and problem-focused discussions *later in the match* all predicted less misconduct at discharge. However, high levels of play *at the start of the relationship* was associated with more misconduct at discharge. These findings suggest the timing of activities is as important as the type of activities the matches engage in.

Second, there were direct, positive effects of being matched with Advocates who had been teachers and with more educated Advocates. Advocates with prior teaching spent less time learning about the youth through casual conversation and engaged in less problem-focused discussion later in their relationship. This appears to be one way that Advocates with teaching experience indirectly affected lower rates of misconduct at discharge.

Conversely, despite the positive, direct effect of having a more educated Advocate, there also was an indirect, negative effect of being matched with Advocates with more education. The negative effect on youth matched with more educated Advocates appears to have resulted from their lower frequency of play later in their relationship. While it is not clear what accounts for the direct, positive effects of these two Advocate characteristics, the findings suggest that hiring ex-teachers and more educated adults to be Advocates may be a useful staffing strategy.

Role of activities and Advocate characteristics. Contrary to our expectations and first hypothesis, activities engaged in initially were not related to the level of misconduct the youth reported at entry. Youths' self-reported misconduct (at entry) did not seem to influence what type of activity occurred first in the match. Rather, initial misconduct rates were better predictor of activities in the second half of the relationship. Thus it does not seem that Advocates with youth reporting more misconduct at entry chose at the outset to engage in more problem-focused (less playful) activities with their mentees.

Problem focused interactions. The results of factor-based assessments of activities revealed the potentially damaging consequences of Advocates' focusing more on their youth's problems later in the relationship. Such Advocate behaviors include directing the youth's attention to what the Advocate thinks the youth is doing wrong or needs to do differently to be successful in the future. This approach is consistent with Morrow and Styles' (1995) *prescriptive mentoring style* and appears to have contributed to higher rates of misconduct reported at discharge.

Problem-discussion frequency in the first half of the match was unrelated to misconduct at discharge, but higher rates of problem-focused discussions later in the match predicted higher misconduct at discharge. This suggests that it may be unwise for a mentor to impose a problem-focus discussion focus after getting to know the youth. Delaying problem-focused discussion until after the mentor has gotten to know the youth may affirm for the youth that he or she is a problem to be dealt with (as confirmed by the mentor's focus) rather than someone adults want to spend time playing with, learning more about (talking with), or helping somehow through engagement in the youth's community (doing with), which are the other activity types in the TEAM framework.

Therefore, attending to when a problem-focused approach is taken in a match may be very helpful in maximizing the benefits of mentoring this type of youth. These findings do not support the use of what Morrow and Styles' called the *developmental style*. It does not appear best, when mentoring court-referred youth, to emphasize cultivating a friendship first and allow problems to arise naturally over time. The findings regarding the role of play early in the match concur with this view.

Playful interactions. The models using both observed scores and factor scores of the frequency of play revealed that matches with a higher frequency of play in the second part of their relationship tended to include the youth reporting the lowest rates of misconduct at discharge, holding constant the influence of youth and Advocate characteristics. This finding held both for observed score analyses, in which all four play indicators made equal contributions to the mean score for play, and the factor score analyses, in which the play assessment was primarily an indicator of the frequency of time spent playing sports.

Playing early in the match did not appear to be a successful approach. Although the same findings was not observed in the path model using factor scores for play, the path models using observed (mean) scores for play frequency found that higher rates of play at the start of the match were associated with higher misconduct at discharge. This finding comports with the interpretation above that focusing on play early and problems later may be less effective than dealing with the problems facing the youth from the start and allowing a friendship to develop through playful interactions later.

An alternative explanation could be that it was youth characteristics (age, sex, or starting misconduct) that explain why early play contributes to higher rates of misconduct over time, except that youth characteristics were only minimally useful in explaining what types of activities matches engaged in during the match. Problem-focused interactions occurred more frequently at both time points among girls than boys, and play occurred less for girls than boys in the first half of the relationship. But neither age or sex of the youth were directly related to rates of misconduct at discharge.

Youth characteristics were, however, predictors of with whom they were matched. Girls' Advocates were more educated than boys' Advocates, and boys' Advocates were more likely to have had teaching experience than girls' Advocates. Older

youth tended to have both more educated Advocates and Advocates with teaching experience. Through these associations, the youths' age and sex had indirect effects on outcomes through influencing with whom they were matched.

Advocate contributions. The level of Advocates' education and prior teaching experience were two characteristics that partly explained what happened in the match over time. We chose these two characteristics because they linked directly to the TEAM framework guiding our analyses of playing, casual conversations, and problem-oriented discussions. The TEAM Framework, described more fully in the full write-up of Study 2, presents the common adult-youth roles of counselor, teacher, supervisor, "vice principal" and coach. It proposes what the youth mentoring literature suggests *mentoring* happens when these conventional roles shift from standard, hierarchical relationships to more reciprocal relationships that balance structure and relationship development. Those propositions were used to inform the hypotheses that were tested in Study 2.

These Advocate characteristics also represent dissimilarity between the youth and the Advocates, which, according to the YAP logic model, should lead to less, not more improvement from program participation. The YAP logic model proposes that youth-Advocate similarity is a key factor in what makes YAP participation most effective. We assumed these two Advocate characteristics—prior experience in the common adult-youth role of teacher and the Advocate's education level—would allow inferences about the role or consequences of youth-perceived dissimilarity between their Advocates and themselves, assuming that the more educated the Advocates were the more dissimilar their youth would perceive them.

Using factor score assessments of activity frequency, we found that those Advocates who had prior teaching experience were less likely to engage in more problem-focused interactions in the second half of the relationship, thus having an indirect positive effect on outcomes through their tendency to engage in fewer prescriptive, problem-focused conversations later in the match. Having an Advocate with teaching experience also had a direct, positive effect on youth outcomes.

The second Advocate characteristic that contributed to the benefits of youths' participation in YAP was the Advocate's level of education. YAP views similarity between youth and their Advocates as a primary way in which YAP Advocates are able to reach youth and help them. Assuming most of the youth in YAP are from families with lower than average educational attainment, which may be incorrect, Advocates with college degrees may be expected to be less effective by virtue of their educational dissimilarity from their mentees and their mentees' family members.

We found a direct positive effect of Advocates' levels of educational attainment on youth outcomes. Those youth mentored by more educated Advocates engaged in less misconduct at the conclusion of their time in YAP. Although this association, after accounting for the youths' age and sex.

There was, however, a potentially negative, indirect effect of Advocates'

educational attainment in that matches with more educated Advocates were less likely to spend time playing later in the relationship. Because the amount of time Advocates played with their youth later in the relationship was a strong predictor of lower reports of misconduct at discharge, this may be one way Advocates' education was not helpful.

The sex of the youth played an important moderating role. Lower levels of play were reported in matches with more educated Advocates, and older girls were most likely to be matched with the more educated Advocates, so it could be that older girls are simply less likely to engage in sports-type play with their Advocates because it is less developmentally "appropriate" or gender-role incongruent. Yet the absence of a direct relationship between age and rates of playful activities does not support this view.

These findings point to ways in which Advocates could be recruited, trained and supervised differently. Recruiting more educated Advocates and those with teaching experience may provide one way to boost program impact, even though it conflicts with YAP's logical model which places a premium on hiring Advocates who are most like those they serve. But, regardless of educational background of the Advocates hired, training could be developed to help increase the Advocates' appreciation for the role of play and positive experiences in the match and for helping them to be wary of taking on an overly prescriptive, problem-focused approach later in the relationship. Once Advocates are trained, staff could provide ongoing supervision and encouragement that helps those least likely to play (such as the more educated Advocates) consider its benefits and the reasons for engaging in more play over time in their matches. Finally, trying to determine why having prior teaching experience may make those Advocates more effective could inform both the training and supervision of Advocates regardless of their backgrounds as well.

II. The Four Agencies

Youth Advocate Programs, Inc.

Youth Advocate Programs, Inc. (YAP) is a national leader in community-based services for young people involved in the juvenile justice, child welfare, and mental health systems. Based in Harrisburg, PA, YAP currently operates programs in twenty states, serving over 2,000 young people and their families at any time. The organizational mission is to provide individuals who are, have been, or may become system involved or subject to compulsory care with the opportunity to develop, contribute, and be valued as assets so that communities have safe, effective, and efficient alternatives to institutional placement.

YAP is distinctive in multiple respects, most notably for its focus on non-residential community-based programming targeting youth at immediate risk of institutionalization due to violent or repeat property offenses. This targeting of youth who committed delinquent acts means that YAP engages a uniquely challenging youth population otherwise not served by mentoring programs. This form of "reverse cherry-picking" is an intended consequence of YAP's "no eject – no reject" policy. By this policy to accept all referrals from juvenile justice and child welfare agencies, the sample included in this study reflects both a difficult population but also one perhaps more representative of the true population of youth who are adjudicated and for whom effective interventions are needed.

A consequence of YAP serving a broader and more representative population of adjudicated youth who had committed delinquent acts makes it difficult to identify an accurate comparison group from non-residential settings, and even to compare the magnitude of outcome or change estimates with those of other programs that may restrict their program to a less challenging population. For this reason, it was critical that the methodology used to assess relative effectiveness or change over time be robust to threats to the validity posed by selection threats.

Beginning at referral, YAP Advocates focus on developing a trusting relationship in order to ensure that the youth satisfy community service and court mandates within the framework of restorative practices. Such concerns are rarely present in delinquency prevention and early onset disruption-focused initiatives. In addition, YAP Advocates aim to strengthen family and community relations in order to create or improve resources that will continue to deter harmful conduct after the period of YAP engagement is over.

YAP Advocates are recruited from the same zip-code as youth and families to ensure that service providers have a grassroots understanding of community assets, challenges, and opportunities. To the extent possible, YAP Advocates are then matched with youth based on their similarity and shared interests. Unlike many adult mentors who volunteer for programs such as Big Brothers/Big Sisters, YAP Advocate mentors are

paid, which is a potentially important variable regarding intensity of service delivery and youth outcomes (Friends of the Children, 2011).

In contrast with most advocacy and mentoring programs, YAP's length of services (*i.e.*, dosage) is an average 4-6 months. This shorter term of duration contrasts with typically prescribed one-year mentoring relationships, but the abbreviated service period theoretically is offset by a greater intensity of mentoring contact between Advocates and youth. In fact, when dosage is considered the number of contact hours, versus the number of days or months of being matched, it can be said that YAP reflects a high-dosage, time-constrained approach to youth mentoring.

Research Sites

Study group data was derived from court referred youth in four YAP programs in separate and diverse regions of the country (Camden, NJ, Las Vegas, NV, Lebanon, PA, and Toledo, OH). Sites were selected according to five main criteria (program size, location, data availability, delinquency severity, and generalizability). While YAP programs vary in the number of youth referred and served, a review of program characteristics indicated that programs serving approximately 50 youth at a time are most representative of YAP programs. Programs in separate geographic regions were selected for generalizability purposes. Selecting programs with legal referring authorities that promised access to data records for outcome analysis was also a key selection factor regarding necessary data availability, as were programs focusing on serving youth representative of the "type" of youth who had repeatedly committed delinquent acts who are most common in YAP programs. In short, site selection sought to include programs both representative of a national scope and reflective of YAP programming. This section of the report provides a descriptive overview of each of the four sites.

Camden, NJ

Referral process. Participation and placement in the Camden YAP reflects a combination of court orders and community-based organizations (e.g., Youth Christian Ministries) and self-referrals. YAP maintains a relationship with the juvenile court system through Community Reintegration into the System (CRIS), a defined initiative for at-risk youth. About 20% of youth clients are technically court ordered into the CRIS program, which is executed through the delivery of YAP services. Much of the balance of the program's population is a function of referrals come from juvenile probation and parole with YAP participation a formal condition.

Youth population characteristics. The local juvenile judge refers lower level, as well as chronic, offenders to YAP – a modality divergence justified by the site Director who is committed to treating all youth as high risk for deterrence purposes. The demographics of program youth were a racial distribution of 85% African-American, 10% Latino, 4% White, and 1% Asian and a gender balance of 75% male and 25 %

female. The majority of clients were 16-19 years of age and from broken homes.

Delinquent behavior reflects heavy involvement in property crime and drug activity. Curiously, most male drug use was limited to marijuana but female youth reported harder drug use. Related, two females interviewed admitted prostitution for drug money. The largest contextualizing factor for the Camden site, however, was gang presence. Program youth are at-risk of gang involvement (Bloods, Crips, MS13) although not many were actually members in them yet.

The Camden YAP youth were the most pronounced of all the sub-samples of youth across study sites in terms of serious delinquency. Most were from Camden proper, many from the Whitman Park area known for violence and open-air drug dealing. The majority of male youth interviewed had witnessed shootings, including a murder in broad daylight, years of gang turf warfare, and drug dealing as normative, and expressed a clear internalization of criminal subculture (e.g., endorsement of the stop snitching movement). The majority had been placed in Camelot and Brook Academy, an alternative school.

Service capacity. The Camden program is large relative to other YAP sites, having served more than 100 youth each year of the study period with 86 active youth at the date of the last site visit and another 80 on a wait list. The program has a contract with the State of New Jersey to serve a minimum of 40 juvenile clients for approximately 400 hours a week. Though desirable, major issues bar expansion including difficulty recruiting additional Advocates and limited resources. Although the Camden facility is housed in a large office building setting with a fair amount of space, planning is ongoing regarding acquiring a stand-alone building (so as to qualify for a tax credit from the City of Camden) to centralize general positive youth development as well as house treatment, advocacy, administration services.

The majority of Advocates identify with the youth they served, having grown up in the same neighborhoods and been exposed to numerous risk factors. Some Advocates were recently certified as substance abuse counselors, and the site had received external funding for substance abuse treatment. All Advocates had completed uniform YAP training, and had either completed or had scheduled additional training – all online. Generally, the Advocates were well versed in YAP procedures and policies, and demonstrated considerable enthusiasm in a spirit of giving back and community service.

Activities. Advocacy with the family thematically characterized the general orientation to services in Camden. Advocates experienced various reactions from the families (typically single, female-headed households), ranging from supportive, to indifferent (i.e., thought to view YAP as free babysitting), and even hostile (per the position that YAP was superseding parental authority). The Camden site also involved parents in YAP activities such as cookouts and wraparound services (e.g., anger management and drug treatment).

The primary activity between Advocates and youth, not just in Camden but also in Toledo, was basketball as confirmed by all Advocates and most youth interviewed. Seemingly innocuous, repeated narratives regarding the value of basketball (e.g., building teamwork and interpersonal skills) highlighted a pronounced distinction between the salience of individualized versus group-based services delivery. Additional activities included a book club where youth were able to meet some of the authors of a book they had read and, when required by probation, community service. YAP youth court ordered to perform community service did so through participation in Youth Build. Advocates had also taken youth to sporting events, museums, movies, and public parks during their weekly meetings with specific activities determined according to an individual youth's interests.

Program fidelity. The Camden program demonstrated program fidelity in terms of exposure/dosage, stakeholder engagement, and (lack of) program differentiation. The office projects professionalism and all records were in order. The program as delivered, however, does not strictly adhere in all respects to the prescribed YAP modality. Whereas the YAP modality prescribes that match activities should often transpire in a one-to-one context, budgetary constraints resulted in a predominance of group-based time.

Las Vegas, NV

Referral process. YAP collaborates with the Las Vegas Metropolitan Police Department Vice Unit, juvenile court, and juvenile probation to identify potential youth for referral into the program. The program is court ordered and participation is mandatory, typically as a requirement of juvenile probation. Judge William Voy presides over the juvenile court docket for Clark County and devotes one day per week to dealing with youth referred for prostitution-related charges. In particular, the Wednesday morning court calendar deals exclusively with these types of youthful offenders. Judge Voy is a frequent source of referrals to YAP, as are the juvenile probation and juvenile justice systems in Clark County. At this particular site, the program has the full cooperation of referring authorities, leading to high numbers of enrollment.

Youth population characteristics. The Las Vegas YAP services both male and female youth between the ages of 11 and 18 years. In terms of racial/ethnic background, the Las Vegas YAP was diverse, with White, Black, Hispanic, and Asian/Pacific Islander youth represented. Most of the program's male participants hailed from low socioeconomic backgrounds while there was more class diversity for females. More specifically, female participants were from a broader range of SES backgrounds, including those from the middle classes. Unlike the other YAP sites, the Las Vegas program is unique in that the majority of youth serviced are females involved in commercial sexual exploitation. Conversely, common offenses leading to referrals for males include property crime, graffiti, vandalism, and breaking and entering. In fact, Las

Vegas has one of the highest rates for juvenile breaking and entering nationwide, a reality reflected in the study's sample.

The overwhelming majority of females (90%) served by the Las Vegas program have a history of commercial sexual exploitation (CSE). Evidence suggests that about half of these youth are from Las Vegas while the other half are brought from outside of Nevada. Often these are from neighboring Western states such as California or Oregon, but there have also been instances of youth brought from across the U.S. as well as internationally. The director of the program relayed that more of these referrals appeared to be "homegrown" in recent years. The YAP program in Las Vegas is one of very few that serves this particularly vulnerable class of juvenile offenders.

Service capacity. Las Vegas YAP has multiple contracts with the Clark County Department of Juvenile Justice System to service various types of youth. For the traditional YAP advocacy model, the site has a contract to serve up to 23 youth annually. The Las Vegas site also has a grant award from the Department of Juvenile Justice to assist 10 youth with employment support training and services. The Las Vegas site is among the largest YAP programs evaluated and has an average of 10 youth on the waiting list at any time.

Per the YAP modality, each youth receives approximately 30 hours per month which translates to 7.5 hours per week. Advocates and youth are free to design their own schedules and split up their hours as they see fit. Most schedules depend on both Advocates and youth commitments each week.

Training for Advocates at the Las Vegas site involves both the typical YAP-specific training (basic and continuing), but also requires additional training for their Advocates because of the CSE element. This site also was more successful in terms of demographic matching between Advocates and youth. For example, 50% of the staff is bilingual (Spanish speakers) and zip code recruitment is common for identifying potential mentors for youth. Advocate retention was high at this site relative to the others and the Director used her position as an adjunct professor in the Department of Criminal Justice at UNLV to maintain a steady pool of potential applicants. Each Advocate is part-time at this site (20-25 hours per week) which is intentional on the part of the Director. Advocate burn out appears to be less of an issue for part-time employees. Matches between Advocates and youth typically last until youth are released from probation (about 3-6 months per match).

Activities. As noted above, each youth receives 30 hours per month of advocacy services. Half of these hours are individual one-on-one time between youth and Advocate based upon the youth's Individual Treatment Plan (ITP). This plan is done at the start of the mentoring relationships, and includes a lists objectives and goals for each area in which the youth requires assistance. The ITP also includes a timeframe for each of the goals and objectives, and the plan is updated throughout the program according to these benchmarks. The remaining half of hours are completed in a group-based format

designed to involve all youth regardless of background or specific needs, such as assisting at homeless shelters.

Most activities at the Las Vegas YAP are gender-segregated. At both site visits, males and females were not observed participating in activities together, and there is little overall interaction between the genders. The Director estimated that for every need for which males require address, there are three needs for which females require services. For example, at the first year's site visit, 70% in 'Girls Circle' group reported being raped during an observed session, a reality not experienced by the male participants. Males participate in 'Boys Council', which serves to address their developmental and criminogenic needs.

Because Las Vegas has one of the worst school dropout rates in country, this program emphasizes school-related activities and assistance (tutoring, acting as a liaison between youth and school, etc.). Other activities include the aforementioned "Girls Circle" and "Boys Council", anger management classes, field trips to cultivate individual interests (e.g., trips to museums), parenting classes (for those youth who are pregnant or who already have children), and visits to the mall, restaurants, or movies.

Program fidelity. The Las Vegas YAP program adhered closely to the design of the modality, as indicated by Advocate training, program components, prescribed caseload, intake timeliness, and the development of individualized treatment plans. A greater breadth of programming and activities were offered by this site and all Advocates were highly qualified, with all possessing at least a Bachelor's degree. This site more than others adhered to demographic matching between youth and Advocates, even going so far as to actually execute zip code matching – a reality not common at the other sites due to logistical issues.

This site also did well in terms of meeting dosage requirements for youth. While few other sites utilize true one-on-one time between Advocates and youth, Las Vegas YAP was so well staffed and organized that at least half of all interaction between mentor Advocates and their youth mentees were individualized. Moreover, the activities engaged in group format (i.e., Girls Circle, Boys Council) were more robust and treatment oriented that group activities at other sites (i.e., playing basketball). Participant engagement was also high at this site relative to others, although this was the case more so for females than males. Universally across the sites females appeared to be more engaged than their male counterparts. This modality may work better with female offenders since they are usually more amenable to the social bond and appeared to enjoy the camaraderie of the group activities much more so than has been observed with the males.

Finally, this site did well with respect to program differentiation, as indicated by fluctuations in program size (the program size was stable), program budget (this site was able to obtain outside funding for additional activities), Advocate caseload (the Director made sure not overload Advocates), and continuity of staffing (both the Director and

most of the Advocates were the same across both site visits).

Lebanon, PA

Referral process. The Lebanon YAP site is populated from two primary referral sources – the County's Department of Children and Youth Services and the County's Juvenile Probation Office – with about 55% and 45% of program participants coming from each source, respectively. Court referred youth typically participated as a condition of probation. Juvenile probation facilitates YAP placement with the understanding that YAP is a resource for addressing chronic delinquency.

Youth population characteristics. Many of the youth participating in the Lebanon YAP had engaged in delinquent acts less frequently and had engaged in less severe delinquent behavior than at the other YAP sites. The only rural site setting in the study, common delinquent behaviors included truancy, drugs, vandalism, and shoplifting. Though less frequent, more serious criminal behavior among Lebanon YAP youth included property crime, sex offenses (including sexual assaults), arson, school violence, and weapons at school (mainly knives). Drug and property offenses accounted for approximately 30% and 25% of off charges, respectively. The second year site visit to Lebanon revealed more serious behaviors in the histories of youth referred to YAP. These histories included gun possession and robberies. Truancy also became a more common reason for referral to YAP at least in part because the definition of chronic truancy – one type of dependency – had been reduced to 10 unexcused absences (the majority of YAP youth had about 25 such absences.

As rural Pennsylvania became more diverse, YAP received more referrals of young people who were gang involved. These population changes were reflected in the fact that 50% of YAP Lebanon participants were Hispanic, followed by 45% White, with a balance of African-American, Native American, and mixed race youth ranging between 13 and 17 years of age.

Service capacity. The Lebanon YAP site serviced an average of 100 youth a year during the evaluation period with approximately 60 probation office referrals and 40 referrals from Children &Youth Services each assessed year (loosely served in two roughly equal cohorts in six month durations). The site employs ten Advocates (five males and five females), four of which are fulltime and, like all other YAP sites examined, makes same gender and same zip-code matches. Though the site has four bilingual Advocates, the swell of Hispanic clients and the likelihood of more in the future pronounced the need for additional Spanish speaking mentors. Part-time Advocates work with three or four youth at a time while fulltime Advocates work with seven to ten youth on average. Advocate recruitment was not found to be a problem as with the Camden site due to successful recruitment efforts at nearby Lebanon Valley College, through Career Link, and word of mouth. All Advocates were initially trained per YAP' uniform national training and continued training was delivered by Rutgers University School of

Social Work. The site had experienced little trouble with Advocate turnover, thereby minimizing program differentiation concerns.

Activities. The Lebanon site delivered the traditional YAP juvenile justice modality in a highly structured and professional manner. Activities reported by Advocates and youth included visits to a local YMCA, fossil hunting, community service (particularly volunteerism at a local foodbank), movies, and help with homework. Due to the aforementioned tightening of the definition of truancy and related to homework, many YAP participants had been placed in cyber-school – an unsuccessful alternative according to the Director, Advocates, and the youth (some of whom did not have access to computers and/or the Internet in their homes or elsewhere outside of YAP).

While the Lebanon site, along with the Las Vegas site, demonstrated adherence to the modality theory, generally and, especially in regard to the delivery of individualized treatment, Advocates also involved youth in organizational group activities ranging from Christmas parties, a community garden planted and maintained by YAP, athletic events, and cookouts with the youths' families.

Program fidelity. The Lebanon site demonstrated a high level of program fidelity, second only to the Las Vegas site. The program featured fairly strict adherence to with the YAP modality design, met or exceeded dosage/exposure expectations, and demonstrated quality delivery of services. The participants were highly engaged in programmatic activities and there was very little program differentiation. Unfortunately and despite a letter of understanding promising the availability of outcome data, no one at the site or the referring agency was capable of synthesizing and transmitting performance measures data – especially disappointing in that the research team provided each site a spreadsheet with variable names and operationalization for recording and transmittal.

Toledo, OH

Referral process. The Toledo YAP program receives referrals from Lucas County Juvenile Court pursuant to a contract between the County and YAP. The agreement was to serve up to 20 youth at a time, a number that is maintained throughout the year. Youth are court-ordered to participate, typically as a condition of their probation. Intake and assessment is done within 48 hours of the referral and matches are made shortly thereafter.

Youth population characteristics. Youth at this site were overwhelmingly male (>90%) and African American (>90%). Females only make up about 20% of court-involved youth in Ohio and tend to be referred to services other than YAP by the juvenile courts. Of the youth served at this site, some were charged with misdemeanors but many had either been charged with felonies, including weapons and drug offenses, or scored moderate to high risk on the Ohio Youth Assessment Scale. The most common types of offenses committed by YAP youth include breaking and entering (the most common offense), drug use (usually in the context of probation revocation) or possession, and

assault. Notably, this site serves youth involved in gangs although these groups tend to be more localized (*i.e.*, "cliques" from housing projects) than national (*e.g.*, Blood, Crips). The Director reported that almost all youth referred to YAP are gang-involved. This site serves youth between the ages of 10 and 18, although most are between the ages of 14 and 17.

Service capacity. The Toledo site was operated under the direction of the Director who had a multi-year career in mental health treatment delivery services and had previously worked with case management for children and families. The Director herself had been an Advocate at one time which contributed to operation of the site. This site employed 14 Advocates for a reasonable 1:4 Advocate to youth ratio, although the caseload does vary depending on the number of youth enrolled at any time. Generally, this program adheres to gender matching although there are times when female Advocates are matched with male youth due to the overwhelming number of males referred relative to females. When female Advocates are paired with male youth, a strength-based matching approach is used.

Per the YAP modality and the contract with the county, each youth receives approximately 10 hours per week (more than the typical 7.5 hours per week established by the modality). Youth may receive up to 10 hours per week, typically split into six hours of individual interaction and four additional hours of group activity. Advocates and youth are free to design their own schedules and split up their hours as they see fit. Like other sites, most schedules depend on both Advocates and youth commitments each week. Much of the effort made on the part of the Advocates involves going into the youth's community, picking up them up, and transporting them to where they need to be (usually--change to some of which is probation-related requirements).

Training for Advocates at the Toledo site involves the typical YAP-specific training (basic and continuing). Recruitment of Advocates includes job fairs, local colleges, and referrals from current Advocates. The Toledo site also strives to adhere to the zip code matching policy of YAP and is fairly successful in their efforts. The Director described Toledo as a relatively small community where people know one another; consequently, it is easier to find suitable Advocates in a setting such as this.

Activities. Activities at the Toledo site were similar to those observed and identified at other sites. These include trips to the library, social activities with other YAP youth, and visits to Toledo's minor league baseball team's games. This site was one to specifically mention providing assistance to families as well as youth. This site also engages in job preparation and training as unemployment among both the adult and youth population in Toledo remains a pressing issue for the community. Activities observed at the program's location were geared mainly toward this goal of finding and maintaining gainful employment.

Other activities include recreational activities, educational field trips, and transporting youth to and from organized sports practices. The Advocates also assumed a

role as liaison between youth and both probation officers and teachers. Advocates relayed that individualized time was typically devoted to working on youth's ISPs. Finally, many Advocates reported assisting their mentees with school-related activities such as tutoring and homework assistance.

Program fidelity. The Toledo YAP appeared to possess relatively high levels of program fidelity. This program adhered fairly closely to the design of the modality, as indicated by Advocate training, program components, prescribed caseload, intake timeliness, and the development of individualized treatment plans. This site also did a satisfactory job at adhering to demographic matching between youth and Advocates. Like the Las Vegas site, Toledo made considerable effort to identify Advocates from the same communities as the youth they served. In terms of meeting dosage requirements for youth, this site actually exceeded the 7.5 hours per week required by the advocacy modality (i.e., 10 hours per week). Intake and assessment is done within 48 hours of the referral and matches are made shortly thereafter. Participant engagement was also fairly high at this site relative to others, and the research team observed quality interactions between youth and Advocates while at the study site. Overall, youth appeared engaged in activities and expressed confidence in their assigned Advocate. Finally, this site did well with respect to program differentiation, as indicated by fluctuations in program size (the program size was stable), program budget, Advocate caseload (typically not exceeding 4:1 youth to Advocate ratio), and continuity of staffing.

Concluding Comments

This study was intended to provide a detailed review of the degree to which each site implemented the national YAP model with fidelity along several dimensions. In the end, only brief commentary or high-level assessments of program model fidelity in program implementation are provided at the end of each section.

Because the original co-PI on the project, Dr. Mitch Miller, left the project before the final year of the study period, staff and agency administrators in each site never had, to my knowledge, an opportunity to review and discuss his findings regarding degrees, types of fidelity, strengths and weaknesses of model implementation in that site. Furthermore, because I was not given access to the data upon which initial fidelity comments and reviews were based, I have no way to assess the veracity or rigor of the data analyses. For both of these reasons, I do not feel I can present a full picture of what was learned to answer the questions under investigation in this part of the study. As a result, I have not included in this report all of the information I was given.

However, the information in Table 2.1 provides information on what happened at each location and for how many hours. This should allow useful comparisons about what happened in each setting.

Table 2.1
Descriptive information about treatment dosage in days and hours of contact

		n	Mean	Std. Dev.	Min	Max
Number of **hours** Advocates reported meeting with youth each week at 2 months	Toledo	29	9.33	2.03	1	10
	Camden	14	15.57	4.89	10	22
	Lebanon	29	7.78	3.13	3	22
	Las Vegas	55	7.15	.55	6	10
	Total	127	8.72	3.50	1	22
Number of **days** Advocates reported meeting with youth each week at 2 months	Toledo	25	2.86	.64	1	4
	Camden	12	3.58	1.38	2	7
	Lebanon	24	2.90	1.10	1	7
	Las Vegas	53	2.81	.52	2	5
	Total	114	2.92	.84	1	7
Number of **hours** youth reported meeting with their Advocates each week at 2 months	Toledo	20	8.90	3.6	2	20
	Camden	10	16.00	4.53	5	20
	Lebanon	15	6.35	2.34	1.75	10
	Las Vegas	45	7.86	3.58	1.50	30
	Total	90	8.74	4.40	1.50	30
Number of **days** youth reported meeting with their Advocates each week at 2 months	Toledo	22	2.91	.55	2	4
	Camden	12	3.17	.58	2	4
	Lebanon	16	3.22	1.70	1	7
	Las Vegas	44	2.85	.439	1.5	4
	Total	94	2.97	.83	1	7
Number of **hours** Advocates reported meeting with youth each week at 4 months	Toledo	20	9.90	.45	8	10
	Camden	13	16.15	3.83	9	22
	Lebanon	18	7.66	1.28	5	10
	Las Vegas	47	7.03	.92	4	10
	Total	98	8.94	3.43	4	22
Number of **days** Advocates reported meeting with youth each week at 4 months	Toledo	16	2.72	.68	2	4
	Camden	13	3.15	.80	2	4
	Lebanon	17	2.74	.66	1	4
	Las Vegas	45	2.73	.58	2	5
	Total	91	2.79	.65	1	5
Number of **hours** youth reported meeting with their Advocates each week at 4 months	Toledo	17	9.65	1.06	6	10
	Camden	12	15.00	4.83	3	20
	Lebanon	14	7.11	2.14	3	10
	Las Vegas	45	7.32	1.12	3	10
	Total	88	8.78	3.39	3	20
Number of **days** youth reported meeting with their Advocates each week at 4 months	Toledo	16	3.16	1.20	2	7
	Camden	13	2.77	.81	1	4
	Lebanon	12	3.17	1.03	2	5
	Las Vegas	41	2.81	.44	2	4
	Total	82	2.93	.80	1	7

Though it does not completely describe the range of implementation factors and mentoring processes at each site, the information provided in this first section should give readers of this report a sufficiently detailed picture of what happens at each site to support their understanding of the analyses, results, and interpretations found in the following two sections.

All of the information provided in the first section of the report above was reviewed by our project liaison at YAP for accuracy and appropriateness. No request to change the content of these descriptions was made by YAP so we feel this section provides a relatively accurate description of these sites, program practices, and the approach to mentoring and advocacy processes at each site during the course of this study.

Descriptive Program Characteristics by Agency

Table 2.1 above provides descriptive information about treatment dosage in terms of number of days and hours of contact between Advocates and youth from the perspective of each. Time spent together was reported after two and four months, and is presented below for each agency.

Advocates typically met with their youth on average three days a week and almost 9 hours per week. Dosage across cities was fairly consistent, except for those at Camden who reported twice that many hours together. Cumulatively, meeting 8 hours a week (or 24 hours a month) puts the number of mentoring hours spent together in just the first four months at 100 face-to-face meeting hours (and twice as many in Camden), which far surpasses the number of hours most programs have their mentors meet with their mentees in a full year of mentoring (or the typical evaluation period in most published studies). So, while the YAP model is a short-term program in terms of the number of weeks, it is highly intensive during this time. Furthermore, the youth and Advocate reports are fairly consistent, providing corroborating support that these numbers are fairly accurate estimates of the dosage of the program.

III. Study 1: Outcome Evaluation—Methods, Measures, Design

For these analyses, the recurrent institutional cycle quasi-experimental (RIC) design was employed to address the question about program-related changes on key outcome measures. The RIC design was chosen after it became clear that the propensity score matching approach was unsuccessful because of insufficient data to properly match youth on variables uniquely characteristic of YAP participants. For example, the court referred to YAP youth a disproportionately high percentage of girls involved in sex trafficking in Las Vegas. Without having this variable in the dataset provided by the state from which to identify matches for the YAP youth made creating a comparable matched sample impossible. After this unsuccessful effort to estimate program-related changes on outcomes using propensity score matching, we turned to the RIC approach.

Statistical analysis of program effects on recidivism, deinstitutionalization, and pro-social activity (e.g., school and employment status) using the recurrent institutional cycle quasi-experimental (RIC) entailed estimating change by comparing program graduates (i.e., treatment group) to counterfactual comparison groups of similarly aged youth who had yet to participate in the YAP program, thereby providing the closest approximation of the counterfactual state of what a YAP participant at a given age would look like had that person actually not participated in YAP.

What this design does not allow, however, is the analysis of differences between the treatment and the untreated comparison group on variables not collected prior to entering the program. This is because where no pre-intervention assessments of phenomenon are available, the intervention group cannot provide a counterfactual estimate of the presence of this phenomenon or outcome for the kind of youth who enter a given program at a given age. For example, recidivism was not a viable outcome variable because the data provided by the state did not allow us the identification of dates of recidivism post initial court involvement.

However, there was considerable other data available assessing a variety of crime and recidivism related phenomena to undertake the RIC analyses. Specifically we relied on the data from YAP to assess the graduates' involvement in crime subsequent to graduation, which was collected prior to treatment for all youth, at discharge, and then 12 months post treatment for program completers to be useful for this purpose and for estimating process-outcome links.

Introduction

In the first part of this section of the report the rationale is provided for using the recurrent institutional cycle[3] quasi-experimental (RIC) design to investigate questions about the effectiveness of the YAP program. An explanation of its strengths, data requirements, and necessary assumptions and rival explanation tests (to address specific validity threats) is provided. This section presents a number of preliminary checks of the data and design that serve as background to the main tests of effectiveness.

The first set of assumptions to be checked through preliminary data analyses regard determining what of the available data could provide useful estimates of the counterfactual condition. Namely, which of the assessments, observations, and other data are appropriate for use, first conceptually, then for this particular population of youth, and finally statistically.

What happens in the program should be related to program goals, and both should be related to the data used for the treatment effectiveness study. Because information on what occurred in each site is described in section one of this report, very little additional information is provided in this section.

To identify the best measures for studying this particular program, each of the measures was examined for random (non-program) variance due to developmental processes (e.g., change due to normative maturational processes) and both site and gender variability. Other analyses of the existing data were undertaken to explore how useful different scales would be for tapping into importance constructs, processes, mental states, or behaviors among the population at hand—the youth in the actual sample. Finally, statistical issues related to normality of data as well as robustness to key validity threats are examined.

The Recurrent Institutional Cycle Design

One problem of using experimental designs in program evaluation is the cost of conducting randomized controlled trials of program effects, which are considered by many to be the gold standard of evidence and the only viable means of identifying causal processes. Others, in particular individuals of the very philanthropic foundations that Advocate for the necessity of evidence of program impacts, have initiated explorations of the viability of other, quasi-experimental approaches. But the approach used in this study has been given very little use despite its potential, particularly for widespread use in the field of program evaluation.

Those who only support a true experiment, in which program participants are randomly assigned either to receive or to be denied program services, present agencies like YAP with ethical and moral dilemmas to address. To deny youth access to program services, which they are viewed as needing, in order to undertake a scientific evaluation

[3] Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Dallas, TX: Houghton Mifflin Company.

of outcomes, is neither something that a university research ethics committees will allow nor something that youth's parents readily agree to without some stipulation of promised delayed services. For example, in one novel study of a youth intervention, a school district would not allow a no-treatment, randomly assigned experimental control group. The school district demanded that the researcher provide an alternative service, which was then called a "Lunch Buddy," to students not receiving the intervention under investigation[4]. Much to the researchers' surprise, the comparison condition actually outperformed the focal intervention in specific school settings.

To assess the effect of mentoring program participation for all participants and across age levels, we implement the "recurrent institutional cycle design." This is typically a version of an age-cohort matched design approach; but it goes considerably further in addressing program change through a variety of complementary analyses that rule out rival alternative hypotheses the limit the use of each type of test when conducted singly. In addition, it capitalizes on the advantages offered by recently developed statistical advances that allow statistical dependence to be accounted for and that allow evaluators to exploit the nature of the types of data already collected by many agencies repeatedly each year or across intervention cycles.

The beauty of the design is the careful use of pre- and post-data from youth who participate in the program to provide evidence of post-treatment outcomes and as a highly approximate counterfactual comparison group. Without the use of external comparison groups, the approach provides a uniquely feasible and yet potentially very rigorous quasi-experimental design.

Intervention-duration based cohort creation. Considering the RIC as a variant of the age-cohort quasi-experimental design is a useful way to begin to understand why and how it uses participants' pre and post scores to estimate change associated with program participation. Imagine YAP participants are classified into narrow age cohorts based on the age when they entered the program and completed their pre-test on outcome variables. These youth data reflect their reports on outcome variables at a time before initiating participation in the program. For example, the youth who was 13 years and 1 month at entry (at pre-test) provides one observation of how non-treated youth look on those variables at that age—this pre-test data provides one observation in the sample of the untreated comparison group. This pre-intervention survey is one "picture of" a YAP-type kid at age 13 and 1 month (13:1) who has not received YAP services.

If this youth then participated in YAP for 4 months, he would be 13 years and 5 months (13:5) old when he completed his post-assessment. The post-assessments from

[4] Cavell, T. A., & Hughes, J. N. (2000). Secondary prevention as context for assessing change processes in aggressive children. *Journal of School Psychology*, *38*(3), 199–235.

this youth, then, represent what the kind of youth at this age who participates in YAP looks like on these outcomes after participating in the program.

This is how a given individual provides both information for the treated group and the untreated (comparison) group by being assigned to two statistically distinct age cohort groups--both as a 13 year, 1 month old untreated comparison group member and a 13 year 5 month old treatment group member. Of course, there are many problems posed by this design, but these can be systematically addressed with adequate data.

Design for estimating change associated with program participation. This approach was used to construct a counterfactual comparison group against which to compare participating youths' outcome scores after 4 months of participation in YAP. The pre-test data for entering program participants was chosen to represent the counterfactual state for youth who entered the program at that age.

The second element of this design is that it requires at least three cohorts of participants and the separation of individuals' pre and post-tests data into separate cohorts within each intervention cycle. By making comparisons across cohorts, multiple tests of program-related change can be replicated. By creating within-cohort cycle-length groups addresses fundamental limitations of between-group tests with statistically dependent data. The dependency would otherwise result from the separation of pre and post-data from the individual youth and assignment of each to separate non-independent groups—treated and the counterfactual.

The issue statistical dependency dealt with in two ways. First, it is addressed through the use of within-cohort tests across smaller statistically independent age-cohorts. Figure 3.1 illustrates this quasi-experimental design whereby the pre-test data from youth entering one intervention cycle are assigned to the "untreated" cohort group for their intervention cycle. In that illustration, Lou's pre-test data for intervention cycle group 2 is used as the untreated comparison group that is compared to the "treated" data from Tom's intervention cycle number 1. In this way, the post-test data from one cohort is compared to the pre-test data from the subsequent intervention cycle cohort. This secures statistical independence (Tom's pre- and post-data are not compared).

This does, however, create a sometimes untenable number of small treatment effect comparison groups and lead to many highly underpowered between group tests that simultaneously inflate Type 1 error rates.

When the statistical analyses described in the next section are beyond the skills of evaluators and there is an external demand for just a few adequately powered statistical tests of the likelihood of the observed between group differences, these smaller age-specific (or intervention cycle duration differentiation) groups can be pooled. Using one or two tests (e.g., the second including analyses of covariance) affords greater statistical power and simultaneously provides a basic way to statistically accounts for data dependency, whole-sample tests (including all youths' pre and post-tests data) can be

made. But these tests require some method of maturational processes to be ruled out since the "untreated" group will, on the whole, be one age or intervention cycle-length younger than the "treated" group. Using age as a covariate is one way to address this, while the other is to analyze the amount of change typically observed across the pre-test scores of different ages.

A final method of assignment is to pool all youth in the study into one test of differences between treated and untreated youth that explicitly accounts for data dependency through the nesting of difference-test groups within each person. This can be particularly helpful in countering the fact that analyses of covariance cannot address or sensibly adjust group-specific scores on covariates like sex, race, and prior misbehavior rates because, in the end these are the same for both the treated and untreated groups (by the fact that it is the same people in each group), so there is no mean difference by which to adjust the groups. Therefore, based on the inclusion of the same youth as treated and non-treated group participants, treatment (mentored) and untreated (comparison) youth will create groups that will not differ on any of these demographic or other risk factor variables, such that analyses of covariance are unnecessary to equate groups, except on age. Covariates are simply included to remove from outcome variance that can be explained by known characteristics and which would otherwise be considered random error in the model, making estimation of differences due to program participation less easy to discriminate effectively. Using a second set of difference analyses to address data dependency, covariates are not needed.

It is this second type of whole-sample probability of between untreated and treated group difference analyses (in which these two data points are nested within person using hierarchical modeling) is most valuable. It requires greater skills in statistical analyses that many agency-specific evaluators may have (or have the resources to seek consultation) and so are not essential for sufficient use of the RIC analyses but which, when used, render the RIC one of the strongest quasi-experimental designs.

The corroboration of estimates of change. What is critical for the full implementation of a RIC design, and what makes the RIC different more rigorous that running either the weaker post-test only between-group difference tests or the pre- post-change tests alone, is the use of pre-post and between-group tests in combination and comparing these estimates to corroborate evidence of change generated by each separate approach. Furthermore, and also required for proper use of the RIC is the comparison of program-related change estimates across multiple cohorts. This provides further evidence of the corroboration of program-related differences.

Figure 3.1 below provides an example of how the use of multiple (minimally three) cohorts to conduct multiple within-person and between-group difference tests can be done. Shown in the figure are two individuals, Tom and Lou's', who enter the program in different cohorts. This figure does not show the third cohort because the third

cohort simply allows a replication of what is shown in the figure for two cohorts. For simplicity and clarity, the figure illustrates how tests of the significance or the magnitude of individual's pre- and post-post assessments are made. Within individual change for Tom compares his pre-treatment (at age 12 years, nine months) to his post-treatment score (at 13 years, 0 months). Pre-post change for Lou's represents the difference between his pre-treatment score (at age 13 years, one month) and his and post-treatment score (at age 13 years, 5 months). Notice, there will be both a degree (or magnitude) of difference and a direction of change (increases or decreases of the post-score to the pre-score). These pre-post change estimates reveal whether there was any change related to program participation. In some cases, such as when the outcome is a phenomenon known to increase over the length of time represented by the age or intervention cycle length used to create cohorts, no change can reflect a treatment effect. Otherwise, no change on variables not likely to change as a result of maturation would suggest no program-related change and therefore no program-related treatment effect.

The second difference estimate of change is the comparison of mean scores for the untreated youth and the treated youth. Between-group (across intervention cohorts) change provides the second critical test of program-related change. In this test, Tom's comparison/untreated cohort's pre-test average is compared to Lou's intervention/treated cohort's mean post-test score to generate a treated vs. untreated comparison of estimated change associated with program participation.

Not illustrated in Figure 3.1 is that with one more intervention cohort, the tests of both within-youth (pre-post) and between untreated and treated groups differences can be examined for corroboration across cohorts. Comparing the comparability of these two types of program-related change can either corroborate (or contradict) the estimates of change observed in analyses run with earlier cohorts.

The inclusion of tests across multiple cohorts and comparisons of the size and direction of differences, both pre-post (within) and treatment/untreated (between) group comparisons, provides a third method of corroboration of evidence. The effect size of this treated vs. untreated difference can be compared to the difference between treated and untreated groups in subsequent (and prior) intervention cohorts, such as Lou's group's post-test mean score being compared to the average pre-score of the intervention cohort that follows his. These comparisons can be conducted for all youth in each cohort, or be stratified by age (or other moderators) when maturational processes are at play, or when there is significant variability across cohorts (or across moderating characteristics of the participants who do and don't tend to complete the program: attriters).

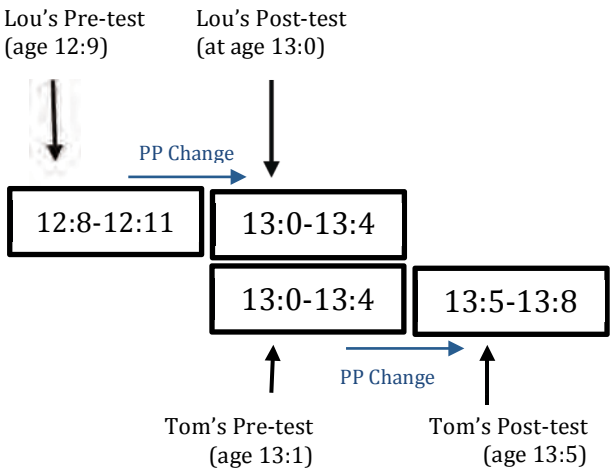


Figure 3.1. Representation of the Age Cohort, Recurrent Institutional Cycle Design

Missing data and attrition. An important issue related to the RIC design's comparison of treated vs. untreated individuals is the problem of attrition and missing data. The comparison of pre- and post-test scores that does not account for attrition is always problematic. Making comparisons across intervention cohort groups without addressing missing data, particularly when non-completion of the program is associated with specific youth characteristics, observed or not observed, will make the counterfactual comparison group much less useful, especially if characteristics related to missing data are also associated with program outcomes. In programs like YAP, the pool of those who complete the program (the treatment group) will often not include the youth who entered the program at greatest risk for misconduct and quit prematurely, such that the comparison group means will likely be less favorable (higher in misconduct) because the highest risk youth are in the comparison group but no longer in the treated group. So detection of missing data and its treatment can be vital to effectively using the RIC design. In this study, FIML was used in the structural modeling of hierarchical (nested dyadic) data and multiple-imputation was used to construct a dataset that estimated scores for individuals missing data at discharge or twelve months post-discharge.

The reliance on standardized effect size estimates. Another way of corroborating evidence is to compare the results of estimated differences between groups (effect size coefficients) to significance tests testing the hypothesis of no difference. Significance tests like those associated with t-tests and F-tests address Type I error (claiming differences that are really due to chance) by relying on the probability of a given difference occurring in a sample of a specific size were this effect not really one found in the population (i.e., by running the same tests with many different samples). Significance tests are most appropriate with larger samples and when fewer (not more)

statistical tests are conducted to answer research questions. Therefore, the nature (tests across smaller subsample cohorts) and number of tests conducted in the RIC design make heavy reliance on significance tests unwise and improper.

In RIC design, the interpretation of difference and change analyses relies heavily on the estimation and comparison of effect sizes. Effect sizes can be more useful in addressing Type II error (failing to identify an effect that is true in the world or larger population) when samples are smaller and when multiple estimates of the same effect can be corroborated. Effect sizes also convey the useful information of how large an effect appears to be and is less adversely affected by small samples than are probability tests.

The effect size used in this study to estimate the magnitude of the treatment vs. comparison group difference in means is Cohen's d. The goal is to estimate how YAP-like youth present on outcomes depending on membership in one of two groups—whether *they have* or *have not* participated in YAP. This difference is estimated using the following equation:

Figure 3.2: Effect size formula used

$$\text{Cohen's } d = \frac{(\text{M}_{\text{Treatment}} - \text{M}_{\text{Comparison}})}{\text{SD}_{\text{pooledl}}}.$$

In this study and design, the primary test statistic used is the standardized group mean difference (see Figure 3.2), Cohen's d, which has specific conventions regarding its interpretation in the social sciences. A d of .20 or larger is considered "small" by convention, but also meaningful in this study. Although .20 is an indication of a "small-sized" effect, it is also the typical program effect size for mentoring programs (as revealed in multiple meta-analyses of youth mentoring studies, see DuBois et al., 2002).

In the results section, there are d (effect size) estimates for both the between-group difference test (treated vs. untreated) and for individuals' own change scores (pre- vs. post scores). These estimates of difference/change are conducted within (pre-/post-test) and across cohorts (pre-/untreated in one cohort v. post/treated in the prior cohort). In addition to within and across cohort estimates, tables in the results section that follow also provide average effect sizes for individual change across time and between-group difference for the whole sample.

Effect size consistency provides a second criterion for assessing the likelihood that effects of program participation are real. In fact, RIC analyses are based on the assumption (especially when intervention cohorts are small) that replication is just as foundational a principle in the scientific method as is probability testing (see Kline, 2012). Finding meaningful ($d \geq .20$) differences in more than one cohort comparison provides greater confidence in these effect sizes reflecting evidence of a true or common effect. Where there are consistent meaningful differences favoring those who have completed YAP on multiple outcomes in a given cohort greater confidence also is

instilled. Yet evidence of consistent differences across measures found for between just two cohorts remains subject to several validity threats, such as history effects for only one cohort's experience that might not be revealed in another cohort. So the use of multiple cross-cohort estimates allows corroboration to determine the validity of this particular threat. Additionally, when comparison across cohorts reveals similarly sized change or difference effect sizes, then the hypothesis that change is due to program participation is even more credible.

It should be noted that there are issues of statistical significance that must be made on a conceptual or philosophical level before proceeding, and that for most of the comparisons made effect size measures of magnitude of an effect will be more reliable that tests of significance. Because there are sub-samples being compared, most tests will be inadequately powered (thus yield a high Type II error rate). But there also are many, many tests being conducted, so reliance on significance tests also could yield high Type I error rates. Although one correction for high Type I error rates is to adjust (by making more stringent) the significance criterion use for test statistics, this is unwise when sample sizes are small as making such adjustments to significance levels is not yet considered an accommodation for small samples sizes by many in the social sciences. Therefore, only the final tests of between-group differences, those with the whole sample (those sufficiently powered by including multiple imputed data sets or the whole sample with assessments nested within youth), consider the need to adjust probability criterion to keep Type I error rates low.

The RIC design depends on effect size estimate consistency rather than significance test probabilities. Of course, if an extremely large sample provides data that allows the smaller cohort sample sizes to be sufficiently powered to be used for the preliminarily validity-check tests described below, then it may be possible to use them and necessary to both adjust the significance level for the number of comparisons (e.g., Bonferroni or Benjamini-Hochberg adjustment) or the size of the sample.

Measurements & Design

Prosocial behaviors and attitudes. This funding initiative was not only interested in recidivism rates subsequent to program participation but also about prosocial behaviors and attitudes. As one phenomenon reflecting prosocial attitudes and behaviors, ecological connectedness captures these elements and has been found to predict criminality and educational success in prior research. These were one set of outcomes assessed as measures of program change.

Hemingway: Measure of Adolescent Connectedness (5.5 version; Karcher, 2003, see Karcher & Sass, 2010). The Hemingway instrument consists of nine scales that assess adolescents' caring for and involvement in specific relationships and activities. The *Connectedness to School* scale focuses on the importance youth place on school and how active they are in being successful in school. *Connectedness to Teachers* assesses effort

to get along with teachers and concerns about earning teachers' respect and trust. *Self-in-the-Future* asks about the behaviors and qualities of youth that they perceive will help them have a positive future. The scales have demonstrated good three-month test-retest reliability, a distinct factor structure, evidence of convergent and discriminant validity, and invariance across sex and race.[5] For the present study, factor scores were computed and fit indices and other analysis output is available by request. Factor scores were used because coefficient alpha reliability estimates for the observed scores for the connectedness subscales when used was *School* (α = .84), *Teachers* (α = .65), *Self-in-the-Future* (α = .82), *Family* (α = .74), and *Friends* (α = .70) were lower than ideal. Factor loadings and fit indices are available from the Author and can be found in the article by Karcher and Sass (2010). The items for each subscale are located in Appendix H.

Items adapted for the Advocate reports of two connectedness scales are presented here:

Connectedness to School

4. This youth works hard at school.
9. This youth enjoys being at school.
14. This youth gets bored at school.
20. This youth currently does well in school.
25. This youth feels good about him/herself at school.

Connectedness to Teachers

6. This youth cares what his/her teachers think of him/her.
11. This youth does not get along with some of his/her teachers.
16. This youth wants to be respected by his/her teachers.
22. This youth tries to get along with his/her teachers.
27. This youth always tries hard to earn his/her teachers' trust.
30. This youth usually likes his/her teachers.

Expects to finish high school, start college, finish college and find a job. From the Big Brothers Big Sisters School Based Mentoring study (Herrera et al. (2007).

"These sentences are about your plans for high school and college. Circle one number to show how sure you are about each question. Check 1 if you're NOT AT ALL SURE or 4 if you are VERY SURE"

Table 3.1 *Expectations for School and Work* "How sure are you that you will"	(Circle one) Not At All Sure	Not Really Sure	Mostly Sure	Very Sure
…finish high school?	1	2	3	4
… go to college?	1	2	3	4
…. finish college?	1	2	3	4
…find a good job?	1	2	3	4

[5] Karcher, M. J. & Sass, D. (2010). A multicultural assessment of adolescent connectedness: Testing measurement invariance across gender and ethnicity. *Journal of Counseling Psychology, 57*(3) 274-289

YAP assessments of educational and vocational engagement. These data are collected regularly as part of YAP's program evaluation

Current living situation. These data are collected regularly as part of YAP's program evaluation.

Table 3.2
Living Situation at Entry

Independently	Foster family	In-patient substance abuse	Homeless
With Parent(s)	Supervised independent living	Non-secure detention	AWOL
Adult Relative	Group home or group residence	Secure detention	
Adult family friend	Residential facility	Incarcerated	
Adoptive family	Mental health facility	Living in a shelter	

Educational (situation) engagement. Only those in bold were used in analyses

Table 3.3 *Current Educational Situation* (If school is currently closed, check the box for when school was last in session.)
Working on or received a Graduate Degree
Working on or received an Undergraduate Degree
Has taken non-degreed Post High School courses
Graduated – received regular high school diploma
Graduated – received GED
Enrolled – attending 4-5 days/week (Home schooled applies)
Enrolled – attending 1-3 days/week (Home schooled applies)
Enrolled but not attending school
Tardy or leaving early at least once a week
Awaiting enrollment in a new school or re-enrollment
Temporarily suspended from school
Permanently expelled from school
Legally withdrawn from school – working on GED
Did not graduate and not working on any type of Diploma
Too young for school / pre-kindergarten

Employment status. The employment status variable ranges from 0 (too young/disability; cannot work), to 1 (unemployed), 2 (actively seeking employment), 3 (working part time), and 4 (working full time).

Self-reported deviancy/misconduct

Misconduct. Below is the format in which the items which were used to create the misconduct measure were collected. These items were originally developed by Bradford Brown et al.[6]

Table 3.4
Self-reported Misconduct

	How often have you done the following in the past year?	*I have never done this*	*Once or twice*	*I did this more than two times*
1.	Run away from home			
2.	Gotten into a gang fight.			
3.	Hit a parent			
4.	Lied to your parents about something important.			
5.	Stole something (worth less than $5)			
6.	Attacked, physically hurt, or "jumped" someone you did not know well			
7.	Took something from a store without paying for it.			
8.	Strong-armed or bullied another kid or student			
9.	Panhandled (asked for money from strangers)			
10.	Strong-armed a teacher (or employer)			

Most serious disposition.

Table 3.5 *Most Serious Disposition at Entry*
None
Non-Criminal and/or Status Offense
Misdemeanor/Citation
Felony

[6] Brown, B. B., Clasen, D. R., & Eicher, S. A. (1986). Perceptions of peer pressure, peer conformity dispositions, and self-reported behavior among adolescents. *Developmental Psychology, 22,* 521–530.

Sample Information

At entry. The sample of the study is described below. At intake, the average age of youth enrolled in YAP and who participated in this study was fifteen and a half years, with a range from eleven to eighteen. More youth were Black (n = 63), but White (n = 44) and Hispanic (n = 39) youth were also well represented in the sample. The large majority of participants were male, with the exception of in Las Vegas, wherein the ratio of boys to girls was 4:3. The number of youth in the Las Vegas site was twice that of youth in Toledo and Lebanon, while the number of youth from the Camden site was a quarter (n = 17) of those from Las Vegas (n = 67).

Data collection for the 164 participating youth in the study ended June 15, 2014. Of these, 164 youth completed pre-surveys, 133 youth and Advocate pairs completed 2-month surveys, and 103 youth and Advocate pairs completed 4-month surveys. To address the issue of attrition, we used multiple imputation to estimate missing data on participant outcomes at the 2-month and 4-month time point for absent follow up data for those who terminated treatment prematurely.

Age. The age at entry did not differ across agencies.

Table 3.6
Age at Enrollment of Participants in Each Site

		N	Mean	Youngest	Oldest
Age at entry (in days)	Toledo	43	15.71	12.00	18.14
	Camden	17	15.96	14.35	18.13
	Lebanon	37	15.40	10.14	17.99
	Las Vegas	67	15.84	11.46	20.37
	Total	164	15.72	10.14	20.37

Sex. The descriptive statistics below are for the whole sample at entry. The proportion of girls to boys was much higher in the Las Vegas site.

Table 3.7
Sex of Participants in Each Site

		City				
		Toledo	Camden	Lebanon	Las Vegas	Total
Sex	Male	37	16	32	38	123
	Female	6	1	5	29	41
Total		43	17	37	67	164

Race. There were significantly more Caucasian/White and Hispanic/Latino youth in Lebanon and Las Vegas than in the other two sites. Las Vegas also had a greater proportion of black/African American youth than Lebanon, but fewer than in Toledo and Camden.

Table 3.8
Race of Participants in Each Site

		City				
		Toledo	Camden	Lebanon	Las Vegas	Total
Race	Caucasian	8	2	16	18	44
	Af. American	30	12	1	20	63
	Hispanic	3	0	17	19	39
	Asian	0	0	0	1	1
	Pacific Islander	0	0	0	1	1
	Bi racial	2	1	3	8	14
	Other	0	2	0	0	2
Total		43	17	37	67	164

Table 3.9 *Presence of Important Adult at Entry*		YAP City				
		Toledo	Camden	Lebanon	Las Vegas	Total
Youth report at entry having an important adult in his or her life	Reported no important adult	14	6	14	33	67
	Reported having an important adult	23	10	18	33	84
Total		37	16	32	66	151

Combined sample characteristics. Below are the characteristics of the sample at entry, discharge, and post-discharge**.** Although some of these descriptive numbers also serve as outcome variables (and thus typically belong in the results section) they are presented here to make clear the differences between the population of youth from whom complete data were provided and for whom missing information at discharge and post-discharge were imputed or addressed with other missing data techniques (in Mplus).

Race/ethnicity.

3.10 *Race/ethnicity at Entry*			
Group Name	Number	Races/Ethnicities Included	%
African American	1	African American	45.9%
Bi-Racial	2	Bi-Racial	4.9%
Caucasian	3	Caucasian	29.4%
Hispanic	4	Hispanic	17.5%
Other	5	American Indian; Asian; Other	2.4%

Primary referral reason.

Table 3.11 *Referral Reason*			
Group Name	Group Number	Referral Reasons Included	%
Legal Issues	1	Involvement w/ the Legal System; Diversion	46.7%
Child Welfare Issues	2	Abuse; Neglect Parent and Child Conflict Parental Support; Domestic Violence Eviction/Homeless; Reunification	24.3%
Mental Health	3	Mental Health	12.9%
School Issues	4	Truancy; Other School Issues	10.8%
Other	5	Substance	5.3%

Living situation. This category reflects responses to the questions youth were asked about their living in the community or whether they had been placed out of home within the prior 30 days. YAP measures living situation at entry or within 30 days prior to entry, depending on whether the youths' referral to YAP was a condition of release from out-of-home placement. The proportions of both groups (exclusions are presented below, with categories ranging from least to most restrictive).

Table 3.12
Living Situation at Entry

Group Name	Group Number	Living Situation at Entry or Within 30 Days Prior to Entry	%
Community Living	1	With parent(s); Adoptive family Adult family friend Independently; Adult relative	78.1%
Foster Care	2	Foster Family	6.9%
Moderate Supervision	3	Group Home; Non-Secure Detention	1.8%
Residential Facility	4	In-patient substance abuse Mental health facility Residential facility	3.5%
Secure Facility	5	Secure detention; Incarcerated	7.9%
Excluded	6	Supervised independent living Homeless, AWOL, Living in a shelter	1.9%

It was unclear whether the 1.9% of youth who were homeless, living in a shelter or listed as AWOL at entry should be included in analyses because YAP typically doesn't have full information on these youth at entry or discharge making it hard to accurately include them in analyses using missing data to address attrition. So this small group of youth was not included in the outcome analyses.

School Attendance.

Table 3.13
School Attendance at Entry

Group Name	Group Number	School Attendance	%
Graduated High School	1	Graduated – received regular high school diploma Has taken non-degreed Post High School courses Working on or received an Graduate Degree Working on or received an Undergraduate Degree	5.0%
Earned GED	2	Graduated – received GED	3.3%
Attending Regularly	3	Enrolled – attending 4-5 days/week Legally withdrawn from school – working on GED	70.4%
Attending Part Time	4	Enrolled – attending 1-3 days/week Tardy or leaving early at least once a week	8.4%
Not Attending	5	Enrolled but not attending school Permanently expelled from school Temporarily suspended from school Did not Graduate and not working on any type of Diploma Awaiting enrollment in a new school or re-enrollment	10.2%
Other	6	Other	2.5%

Employment.

Table 3.14
Employment at Entry

Group Name	Group Number	Employment Situation	%
Employed Full Time	1	Working Full-Time (at least 35 hours a week)	.8%
Employed Part Time	2	Working 20 hours a week or more, but less than full-time Working less than 20 hours a week	3.5%
Actively Seeking Employment	3	Unemployed - Actively seeking employment	17.9%
Unemployed	4	Unemployed - Interested but not looking for work Unemployed - Not interested in working	43.2%
Excluded	5	Too young for most work assignments – under 14 Unable to work due to mental or physical disability	34.7%

At discharge.

Living situation.

Table 3.15
Living Situation at Discharge

Group Name	Group Number	Living Situation	%
Community Living	1	Independently; Adoptive family Adult family friend; Adult relative With parent(s)	84.0%
Foster Care	2	Foster Family	5.3%
Moderate Supervision	3	Group home or group residence Non-secure detention	1.6%
Residential Facility	4	In-patient substance abuse Mental health facility Residential facility	1.9%
Secure Facility	5	Secure detention; Incarcerated	2.4%
Excluded	6	Supervised independent living N/A*; Deceased*; Homeless AWOL*; Living in a shelter	4.8%

School attendance.

Table 3.16
School Attendance at Discharge

Group Name	Group Number	School Attendance	%
Graduated High School	1	Graduated – received regular high school diploma Has taken non-degreed Post High School courses Working on or received a Graduate Degree Working on or received an Undergraduate Degree	5.3%
Earned GED	2	Graduated – received GED	3.0%
Attending Regularly	3	Enrolled – attending 4-5 days/week Legally withdrawn from school – working on GED	73.7%
Attending Part Time	4	Enrolled – attending 1-3 days/week Tardy or leaving early at least once a week	5.4%
Not Attending	5	Enrolled but not attending school Permanently expelled from school Temporarily suspended from school Did not Graduate and not working on any type of Diploma Awaiting enrollment in a new school or re-enrollment	7.8%
Too Young for School	6	Too Young for School	1.5%
N/A	N/A	N/A*	3.3%

* Includes all clients listed as Excluded in Living Situation at Discharge

Employment Situation.

Table 3.17
Employment at Discharge

Group Name	Group Number	Employment Situation	%
Employed Full Time	1	Working Full-Time (at least 35 hours a week)	1.8%
Employed Part Time	2	Working 20 hours a week or more, but less than full-time Working less than 20 hours a week	7.6%
Actively Seeking Employment	3	Unemployed - Actively seeking employment	18.6%
Unemployed	4	Unemployed - Interested but not looking for work Unemployed - Not interested in working	40.6%
Excluded	5	Too young for most work assignments – under 14 or Unable to work due to mental or physical disability	28.3%
N/A	N/A	N/A*	3.3%

At 12 months post discharge

Living situation.

Table 3.18
Living Situation 12 months Post Discharge

Group Name	Group Number	Living Situation	%
Community Living	1	Independently; Adoptive family Adult family friend; Adult relative With parent(s)	88.3%
Foster Care	2	Foster Family	1.8%
Moderate Supervision	3	Group home or group residence Non-secure detention	0.4%
Residential Facility	4	In-patient substance abuse Mental health facility; Residential facility	4.6%
Secure Facility	5	Secure detention; Incarcerated	4.4%
Excluded	6	Supervised independent living N/A*; Deceased*; Homeless AWOL*; Living in a shelter	0.4%

* Note that the remaining questions in post discharge survey may be answered as N/A.

School attendance.

Table 3.19
School Attendance 12 months Post Discharge

Group Name	Group Number	School Attendance	%
Graduated High School	1	Graduated – received regular high school diploma Has taken non-degreed Post High School courses Working on or received a Graduate Degree Working on or received an Undergraduate Degree	7.2%
Earned GED	2	Graduated – received GED	2.9%
Attending Regularly	3	Enrolled – attending 4-5 days/week Legally withdrawn from school – working on GED	79.6%
Attending Part Time	4	Enrolled – attending 1-3 days/week Tardy or leaving early at least once a week	1.1%
Not Attending	5	Enrolled but not attending school Permanently expelled from school Temporarily suspended from school Did not Graduate and not working on any type of Diploma Awaiting enrollment in a new school or re-enrollment	8.0%
Too Young for School	6	Too Young for School	1.2%
N/A	N/A	N/A*	0.0%

* Includes all clients listed as Excluded in Living Situation at Discharge

Employment Situation.

Table 3.20
Employment 12 months Post Discharge

Group Name	Group Number	Employment Situation	% *
Employed Full Time	1	Working Full-Time (at least 35 hours a week)	3.9%
Employed Part Time	2	Working 20 hours a week or more, but less than full-time Working less than 20 hours a week	7.0%
Working at an Internship	6	Volunteer/Internship	0.9%
Actively Seeking Employment	3	Unemployed - Actively seeking employment	7.9%
Unemployed	4	Unemployed - Interested but not looking for work Unemployed - Not interested in working	61.9%
Excluded	5		18.4%
N/A	N/A	N/A*	0.0%

* Includes all clients listed as Excluded in Living Situation at Discharge

Cohort Creation

Typical age-based cohorts did not turn out to be useful and outcome analyses. Nevertheless, a first step was to create cross-cohort treated vs. untreated comparison groups. Subsequently, an intervention cycle cohort grouping was created in which participants were not matched by age by treatment duration as the way to separate individuals across treatment group classifications. A 4-month cohort grouping created a window of time comparable to the average duration of program participation so that the pre-test for one cohort occurs at roughly the same time as the post for the youth in the prior cohort. So, youth who started services within the first four months were placed in cohort 1 (June 1, 2012 – September 30, 2012), the second cohort include youth who started services during the second four-month period (October 2012-January 2012), et cetera.

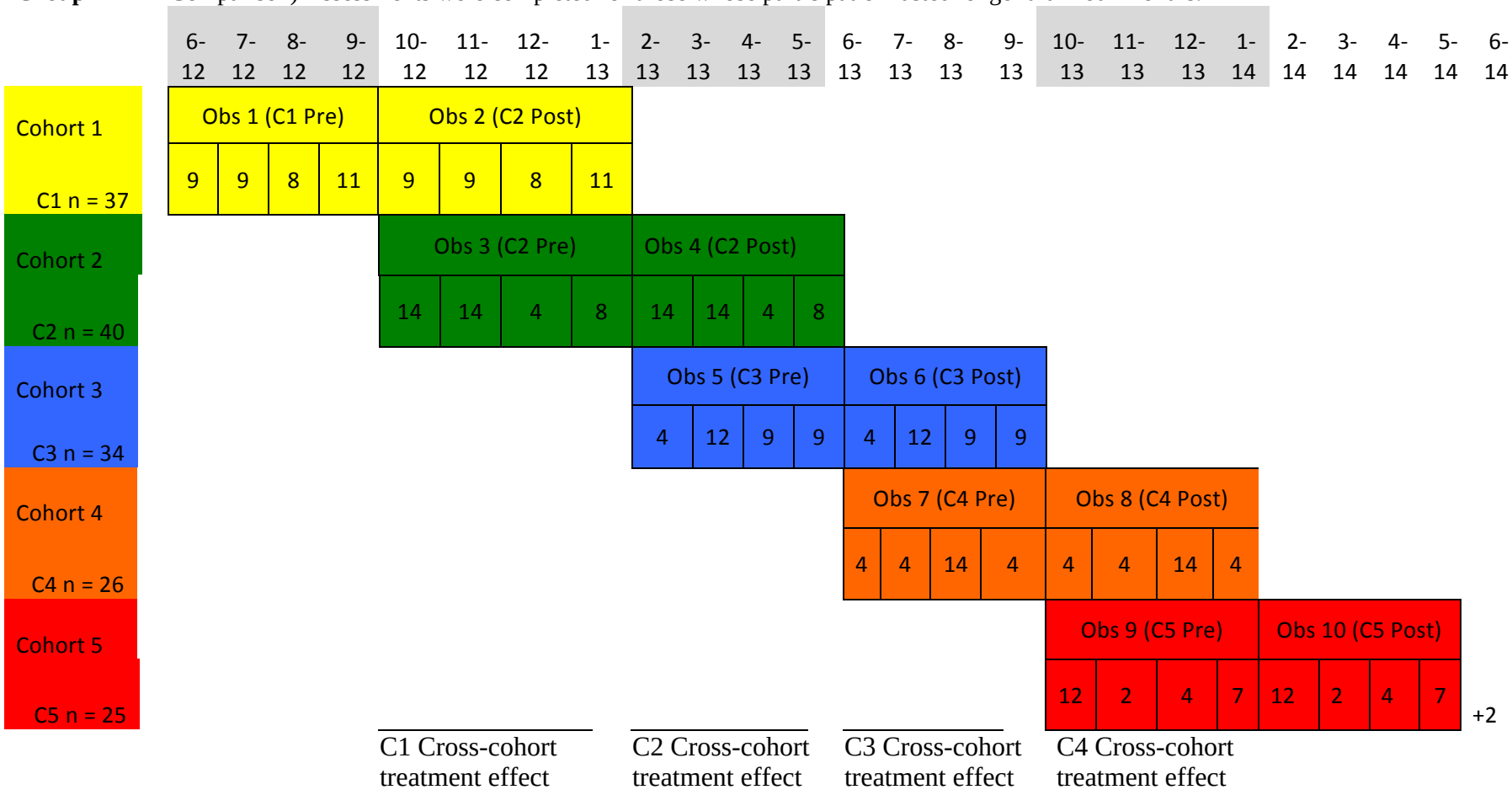
This yielded five cohorts, which ranged in size from 25 to 40 participants. For cohort five, two youth started services in February of 2014, which would have been the first month of what would have been cohort 6. These were the last two youth to be enrolled in the program, as we discontinued enrollment in February of 2014. In order to keep these two youth in the analyses, we included them in cohort five. The 4-month cohort would allow for five pre-post changes comparisons, and four treated vs. untreated tests of difference that compare pre-post scores from the same time period (between cohorts) in which the pre-test is from the subsequent cohort.

The benefit of this cohort grouping is that the length of the cohort inclusion matches the length of treatment we are assessing (4 months) resulting in clean divisions with groupings and testing. Of course, a drawback about this grouping is that the length of time used to group the cohorts' means data are susceptible to history effects. For example, the youth in cohort 1 include those who started receiving YAP services in June through September, and may have variance in their scores associated with the start of school/end of summer or fluctuations in adult supervision associated with being out of school, etc.

This sequence is illustrated below in Figure 3.3.

The Five 4-Month Cohort Group

Period of time, month and year (e.g., October 2012 is 10-12) in which the Post (Treated Group) and Pre (Untreated Comparison) Assessments were completed for those whose participation lasted longer than four months.



*Observations (Obs1, Obs2…) are listed according to original RIC design to help visually identify what data is collected from which cohort when

Figure 3.3 Intervention Cohorts used in Final Analyses

Cohort characteristics. To understand in what ways these cohorts differed at pre-test on key demographic variable as well on the outcome variables a series of comparisons were made. Differences between the cohorts in terms of scores on age, gender, race, city, and on those outcome variables chosen for the RIC pre-post changes and treated vs. untreated comparison t-tests were examined using ANOVA and chi-square tests. An ANOVA was conducted across cohorts as the fixed factor, and age as the dependent variable. For the other three categorical covariates (gender, race, city) chi-square tests were used. For the four main YAP outcome variables (misconduct, serious disposition, education situation, employment status), ANOVA was used.

Age. Results of the ANOVA indicated nearly statistically significant differences between the cohorts in regards to age ($F = 2.41$, $p = .052$). Post-hoc tests revealed the difference in age was between Cohort 2 (15.40, n = 40) and Cohort 3 (16.13, n = 34).

City. Results of the chi-square test indicated that there was a statistically significant difference between Cohorts 4 and both Cohorts 3 and 5 in terms of the number of youth from each City. The difference was largely that Camden had nine youth in Cohort 3, zero youth in Cohort 4, and three youth in Cohort 5.

Table 3.21
Differences in Number of Youth in Cohort 3 & 4 by City

			Cohort Grouped by 4 Mos.		
			3.00	4.00	Total
City	Toledo	Count	9	11	20
		Expected Count	11.3	8.7	20.0
	Camden	Count	9	0	9
		Expected Count	5.1	3.9	9.0
	Lebanon	Count	5	4	9
		Expected Count	5.1	3.9	9.0
	Las Vegas	Count	11	11	22
		Expected Count	12.5	9.5	22.0
Total		Count	34	26	60
		Expected Count	34.0	26.0	60.0

Table 3.22
Chi-Square Tests of Proportional Differences Cohorts 3 and 4 by City

		Value	Df	*p* (2-tailed)
Original data	Pearson Chi-Square	8.39[a]	3	.039
	Likelihood Ratio	11.71	3	.008
	Linear-by-Linear Assoc.	.02	1	.888
	N of Valid Cases	60		

Table 3.23
Differences in Number of Youth in Cohorts 4 and 5 Across Cities

			Cohort Grouped by 4 Mo		
			4.00	5.00	Total
City	Toledo	Count	11	5	16
		Expected Count	7.8	8.2	16.0
	Camden	Count	0	3	3
		Expected Count	1.5	1.5	3.0
	Lebanon	Count	4	11	15
		Expected Count	7.4	7.6	15.0
	Las Vegas	Count	11	8	19
		Expected Count	9.3	9.7	19.0
Total		Count	26	27	53
		Expected Count	26.0	27.0	53.0

Table 3.24
Chi-Square Tests of Proportional Differences Cohorts 4 and 5 by City

		Value	Df	*p* (2-tailed)
Original data	Pearson Chi-Square	8.98[a]	3	.030
	Likelihood Ratio	10.32	3	.016
	Linear-by-Linear Association	.48	1	.488
	N of Valid Cases	53		

a. 2 cells (25.0%) have expected count less than 5. The minimum expected count is 1.47.

Sex and Race. There were no differences in the proportion of either sex or race across the cohorts.

Pre-existing Differences on Variables at Across Cohorts at Pre-Test

Serious disposition. Results of the ANOVA indicated a statistically significant F-test, indicating some differences between the cohorts on serious dispositions at entry. Post-hoc tests revealed that Cohort 5 was significantly different from all the other cohorts in terms of serious disposition. There were no differences between Cohorts 1 through 4. Cohort 5's pre-test data is used as the untreated comparison group in the Cohort 4 (C4) treatment effect test. Because Cohort 5 also was older than the other cohorts, and disposition increases with age, generally, this C4 between group difference test should be given less merit than the other comparisons.

Education situation. Results of the ANOVA indicated there were no statistically significant differences between any of the cohorts on educational status at pre-test. There were differences were between Cohort 2 (2.45, n = 38) and Cohort 3 (2.81, n = 32; p = .008), and between Cohort 4 (2.29, n = 21) and Cohort 5 (2.74, n = 27), p = .09). But the Levine test suggested equality of variance could not be assumed, and therefore the Dunnett T3 post-hoc test was chosen for these analyses. The results of the Dunnett T3 indicated that there were no statistically significant differences between any of the groups, including Cohorts 2 and 3, and Cohorts 4 and 5 once bias related to non-equal variance was accounted for.

Employment status. Results of the ANOVA indicated no statistically significant differences between any of the cohorts in terms of employment status at entry. Initial tests for sequential differences between the cohorts revealed a statistically significant difference between Cohort 3 (1.22, p = 32) and Cohort 4 (1.43, n = 21; p = .011). However, as equality of variances could not be assumed, and the Dunnett T3 post-hoc analyses indicated that there were no statistically significant differences between any of the groups, including Cohorts 3 and 4.

Maturation Tests

Assessing maturation as a threat to program effect interpretations of change is important in the RIC design. To assess the possibility of maturation as an explanation of treatment effects, we examined the degree to which, how consistently, and for what age groups change (observed cross-sectionally) occurs normatively over time on these outcome variables at pre-test. We used the program-prescribed (or average) treatment duration (4 months) as the period across which no observed change on a variable would allow differences observed between pre- and post-test (within person) scores to be considered a function of program participation and not due to normative change over time.

These analyses compare the pre-data for whole sample on mean scores to identify any normative differences on each outcome across age groups differing by 4 months, and where present to assess magnitude (and direction) of normative (cross-sectional)

variability. To make the tests of maturation more liberal (decrease Type II error rates) we choose a timeframe 6 months (or 1.5 times that of the expected treatment duration) for evidence of maturation. Doing this we felt would ensure maturation was not a viable rival explanation of differences found among youth over time: "The observed between-group differences cannot be attributed to maturational processes because change of this magnitude is not normative even over a time period 1.5 times longer than the pre- post assessment interval."

For these maturation tests the grouping of the participants ranged in ages from 14 years 0 months to 17 years 0 months. Too few youth were older or younger than this timeframe to adequately conduct difference tests. These 118 youth, spanning 3 years, were divided into groups based on six-month windows, which yielded seven groups, approximately 1.5 times the typical treatment duration for youth in this study.

The table below lists the number of youth in each group.

Table 3.25
Number of Youth in Each Age Group for Maturation Tests

Age group	14	14.5	15	15.5	16	16.5	17
N	17	13	23	23	14	28	17

A series of ANOVAS was conducted using these age groups as the fixed factor, and the outcome variables of interest (Employment Status, Education Situation, etc.) as the dependent variables (comparing each age group to the subsequent age group: 14 to 14.5, 14.5 to 15, and 15 to 15.5, etc.). We used multiple imputation to estimate the missing data. The average p-value across the five imputation samples pooled was used to infer statistically significant differences across the age groups.

Hemmingway scales. There were no statistically significant differences between the sequential age groups on any of the connectedness scales (friend, family, school, teacher, or future).

Misconduct. There were no statistically significant differences between the age groups in regards to self-reported misconduct in sequential comparisons.

Education expectations: Finish high school. For *Education Expectations: Finish High School*, the difference between age groups 14 and 14.5 years approached a statistically significance ($p = .058$). We considered that because this comparison was between youth who had just started or were about to start high school (14 year olds, many in middle or junior school) and those who had already started (14.5), it made sense that there would be a difference in how youth expectation for finishing high school before and after starting high school. This effect poses more of a history threat to internal validity such that any differences attributed to treatment for youth in these age groups may more accurately reflect this historical event that all of them undergo during this time.

Education expectations: Start college. For *Education Expectations: Start College*, age groups 14 and 14.5 also had a statistically significant difference (p = .03). The explanation above regarding a difference in expectations to finish high school likely applies to this variable as well.

Employment status. For employment status, there were no statistically significant differences between the age groups, except for the groups representing 15.5 (mean = 1.13) and 16 year olds (mean = 1.54), with pooled p-value of .014. For reference, the employment status variable ranges from 0 (too young/disability; cannot work), to 1 (unemployed), 2 (actively seeking employment), 3 (working part time), and 4 (working full time). We felt this difference between the 15.5 and 16 age groups, but not between later age groups, made sense given that the 16 age is typically the first age when youth can legally be employed.

Education situation. For Education Situation, there were no statistically significant differences between the age groups.

Living situation. For Living Situation, there were no statistically significant differences between the age groups.

Serious disposition. For Serious Disposition, there were no statistically significant differences between the age groups; however, the difference between 16 year olds and 16.5 year olds approached statistical significance (p = .07).

In conclusion, there does not appear that for any of these outcome variables the rival threat of maturation is a concern. Very few developmental differences appeared among the many comparisons conducted. On only three variables did differences between age groups suggest the presence of maturational processes. For all these, a difference was only observed between two age groups, and there are very plausible explanations for why two of these differences may occurred suggesting these differences are historical and not maturational. Only for the variable *Serious Disposition* was the reason for the single between-age group difference not apparent. But for this variable as well, developmental differences do not appear to be common or particularly strong and only appeared across on age group difference.

IV. Study 1: Outcome Evaluation—Results

In this section, analyses and findings are reported from most basic analyses first and to most sophisticated or complex analyses last. Some readers (the "statistical purists") may feel that the simplest comparisons are the truest tests of changes associated with program participation. This may be true, particularly for traditional tests of differences across groups when youth are randomly assigned; but for many quasi-experimental group comparisons the more sophisticated and complex analyses often provide better tests of reliable differences, because these analyses are typically run to address more of the possible selection threats and other rival interpretations of findings.

For those findings, readers may wish to start with the final section which reports on tests of group differences that address statistical dependency by nesting data within youth (to create independent statistical tests) and that reflect the greatest statistical power (by being able to pool all youth into one difference test). These comparisons, at the end of this results section, provide direct interpretations, whereas the tests in the first couple of sections require inferring overall change by comparing differences in effect sizes across multiple comparisons.

Then there are the two types of change statistics and two change estimates. The change estimates are the between group differences and the within-person change. Then there are the two types of change-estimating statistics. The first, the effect size, which reflects the standardized mean difference of within and between group changes. The second change statistics is the likelihood or probability test associated with this effect size. The bulk of the analyses in this report are of effect sizes, which are examined for patterns of consistency. This is because we believe that analyzing the consistency of effect sizes and their direction provides an equally persuasive assessment of change. Doing so captures the trustworthiness of estimates of change in terms of their replication. The tables in this section provide both, effect sizes in columns and indicate statistical significance by bolding the effect sizes.

In the final summary section, convergence in the findings across these different analyses is attempted to help the reader integrate the vast number of comparisons made and the several separate tests, each designed to address a specific validity threat.

Preliminary Age-Cohort Matched Comparisons on Outcomes

Recall that understanding how treated and untreated comparison youth of the same age compare on key outcomes is usually a good first step in utilizing the RIC design, even though it alone is not sufficient to make strong inferences. What makes the RIC approach attractive is that it provides a counterfactual estimate of what the average YAP program graduate (treated youth) would look like at that same age had he or she not participated in the program. In the RIC design, the untreated youth are very much like the

treated youth in terms of the typical characteristics of youth who enter YAP, because they are, in fact, the youth who enter YAP. So these cohorts were created in order to run age-matched comparisons.

When within cohort sample sizes are sufficiently large to adequately power the analyses, this is a very useful method that can address chance findings, corroborate through replication, and reveal developmental patterns in program effectiveness.

We attempted to create "treated" and "untreated" groups of youth sharing common 3-month wide age cohorts (at the time of their pre and post-test assessments, respectively). These age cohorts were created to differ by 3 month intervals, which is 30 days less than the typical duration of participation in YAP for the majority of the sample in this study. This would ensure that a youth's pre-test and post-test were included in separate cohorts (3 months apart) so that no youth would be included in the same cohort as both a treated and untreated. That way these comparisons are statistically independent. For example, the pre-score for a "comparison"/untreated youth at age 15 years and 2 months was included in Cohort 8 which included only youth between 15 years 0 months and 15 years 2 months of age. The treatment group youth for this age cohort were program graduates who completed their post-test within this same age window.

Our analyses using these created age-differentiated cohorts generated uninformative results, which had cohorts so small in number that comparisons could not be trusted. For the purpose of running statistically independent tests, cohorts must be created to ensure no individual's pre and post scores are included in the same cohort. These would be very informative had the sample size been larger and in the case that the duration of treatment was longer as well. Because a 3-month age window created 21 cohorts, and because the range in size across cohorts was from 4 to 26, statistical power was simply inadequate for separate tests for each cohort. Therefore these findings are considered most tentative and not very informative. The series of tests that were run comparing untreated and treated youth, just those youth falling within the same cohort at the time of their assessment, on each outcome variable are presented in the appendix, along with a brief overview of the findings is given. Ultimately, these analyses are not very reliable. The results, located in the appendix, are provided mostly to reveal what age groups tended to differ on what outcomes and in what direction for later reference by readers when considering patterns of change over time that may reflect maturational processes.

Changes from Entry to Discharge: The First Reliable Change Estimations

This section reports findings from the tests of within- and between-group differences across cohorts on key outcome variables. First change from entry to discharge are presented and then changes from entry to one year post-discharge are presented. As described above, the tables that present the data provide both pre-post differences within youth and between treated/untreated comparisons across cohorts of

youth. The estimates of change are listed as effect sizes, and the significance of the whole-sample t and F-tests are denoted by bolding.

The tables below share the same format and characteristics and therefore deserve an introduction. First, the outcome variables are listed at the top of all columns but the first, which identifies the specific cohorts (top five rows) or the specific statistical test employed (bottom 3 rows).

Under each variable are two columns. The column on the left includes the effect sizes (*d*) for pre-post (PP) comparisons of individuals own scores within each cohort. There are five rows of pre-post comparisons conducted for the youth in each intervention cycle for cohorts one through five. At the bottom of each PP column is the average pre-post change effect size for the whole sample. Negative effect sizes suggest declines from pre-test to the later assessment (either at post-test/discharge or 12 months post-discharge).

Under the right column (labeled TXCT for Treatment/Control) are four estimates of cross-cohort differences, also reported in effect sizes. The effect size in each row reflects that cohort's post-test scores compared to the subsequent cohort's pre-test scores. As there is no subsequent cohort for cohort 5, there is no Treated vs. Untreated comparison effect size and that cell is blank. Below that empty cell are the between-group (untreated v. treated) differences generated when computed using ANOVA and ANCOVA (controlling for sex, age, race). The bottom row provides the average effect size of the four between-group (cross-cohort) difference tests (from top four rows).

Primary measures of prosocial behavior. The first table provides estimates of change on the measures of connectedness, which served as the primary as indicators of prosocial behavior. Table 2.24 provides the effect sizes for the pre (entry) to post (discharge) assessment points. Descriptive statistics for these tests are in table 2.30 and 2.31. Each of the subtitled sections below summarize these pre-post and treated vs. non-treated comparisons, and reveal the degree to which the separate tests—pre-post (PP), cross-cohort comparisons (TXCT), ANOVA and ANCOVA corroborate each other.

Connectedness to Friends. There was some evidence of program-related changes revealed in the PP and TXCT effect sizes, two of five cohorts demonstrating positive pre-post change and half of the cohorts having higher scores on connectedness to friends among those treated in YAP. The pooled between group difference was meaningful (d = .24). The ANOVA similarly indicated small to moderate increases in connectedness to friends for YAP participants (d = .34); but this relationship was no longer present after controlling for age, sex, race, and city in the ANCOVA (d = .14).

Connectedness to Family. There was little evidence of program effects in the pre-post tests (positive change in only one of five cohorts) and contradictory differences across cohorts (two positive effects, one negative, and one absent any difference). In the ANOVA there were no treated/untreated group differences, but there was evidence of a small positive, statistically significant effect of program participation when controlling for age, sex, race, and city in the ANCOVA (d = .17).

Connectedness to School: The within-person changes from pre- to post changes provided inconsistent evidence, varying across cohorts in both size and direction). ANOVA and ANCOVA also suggested no change in school connectedness.

Connectedness to Teachers: Some evidence of treated vs. untreated group differences was found with the average effect size for cross-cohort tests of change (d = .20) being very similar to the ANOVA tests (d = .22). However, evidence of within person, pre-post change was negligible (only seen in one cohort) and was similar in size to between-group difference found when controlling for age, sex, race, and city in the ANCOVA.

Connectedness to a Self-in-the-Future: Little and inconsistent evidence for change in connectedness to Self-in-the-Future was found in either pre-post or the treated vs. non-treatment tests. Similarly, ANOVA or ANCOVA test showed no evidence of effects.

To summarize, evidence of change between entry and discharge was found on several of the attitudinal and behavioral self-reported prosocial outcomes. The magnitude of the change estimates were typically corroborated across two or more different change estimates. Most strong and consistent was the evidence of improvements in connectedness to teachers. Although pre-post change was observed in just one cohort and between group differences among only two of four cohorts, the average effect size of the four cohort specific tests of treatment related differences was .20 and was consistent with the ANOVA. The multilevel analyses corroborated these findings.

Estimates of pre to post change on the other connectedness scales, was less consistent, with effect sizes either differing in size across cohorts or in direction unexplainably. Consistent pre-post and between group differences were found for Cohort Three across all but Connectedness to the Future. Program graduates in Cohort 4 reported greater connectedness to friends, family, and teachers than youth in the untreated comparison group in the subsequent cohorts. Cohorts 1 and 2 demonstrated almost no change and Cohort 5 had mixed change estimates. Descriptive statistics and test statistics are presented in Table 4.2 for ANOVA and 4.3 for ANCOVA.

Table 4.1
Cohen's d effect Sizes for Pre Post and Treated v. Untreated Difference T-Tests

	Hem Friend		Hem Family		Hem School		Hem Teacher		Hem Future	
Cohort Number	PP	TX CT	PP	TX CT	PP	TX CT	PP	TX CT	PP	TXC T
C1 n = 37	.09	-.02	-.19	**-.27**	-.08	.03	-.14	-.10	-.05	-.03
C2 n = 40	.10	.15	-.02	-.01	.13	.07	.09	.02	-.14	.05
C3 n = 34	**.43**	**.41**	**.34**	**.23**	**.34**	**.45**	**.39**	**.47**	.19	-.04
C4 n = 26	.05	**.45**	-.06	**.27**	**.27**	.11	.18	**.40**	-.08	.19
C5 n = 27	**.26**		.14		**-.24**		.05		**-.27**	
ANOVA d		**.34**		-.04		.10		**.22**		-.11
ANCOVA		.14		**.17**		.10		.11		.13
Pooled PP/TC *d*	.15	**.24**	.04	.06	.08	.17	.11	**.20**	-.07	.04

Notes: PP denotes the effect size for a pre-post t-test comparing the average difference across youths' pre- to post scores within a single cohort. TXCT denotes effect sizes that reflect the magnitude of the difference between 4-month assessments of the treatment group and pre-treatment assessment scores for the subsequent cohort of youth (untreated comparison group). Positive scores indicate increases or higher scores for the treated youth. Below these are pooled sample untreated to treated group differences estimated using ANOVA and using ANCOVA. The bottom row provides the average pre-post difference across all individuals and under TXCT is the average effect size across the four treatment comparisons. All reflect Cohen's d as the effect size, for which d = .20 is "small," and a d= .40 to .50, is "moderate" in size. For the top five rows, bolded scores signify non-negligible (small or d $\geq$ .20) differences. In the bottom three rows, bold reflects statistical significance of $p \leq .05$.

Hierarchical analyses that pool all youth into one between-group test. The above analyses are not subject to bias from data dependency because all youth's pre and post scores were separated by or across cohorts (except for the pre- post tests). But they do not yield a single significance, only effect sizes and (for the ANOVA and ANCOVA) evidence of whether any between-group difference varies across cohorts.

Hierarchical analyses that nest the individuals' pre and post scores can remove or account for the statistical dependency resulting from pooled data and thereby offer a sample size more appropriate for adequately powered significance tests. Using hierarchical analyses (Mplus two level analysis) the Hemingway connectedness scales were compared across treated and untreated comparison groups. The sample size was not sufficiently large to use a probability cutoff score of $p < .05$ to detect the expectably small effect sizes (typical of mentoring, d = .20, see DuBois, et al., 2002), so a p-value of .10 was considered indicative of a non-chance finding and $p < .15$ as a trend.

Given the probability criterion of .10, the treated youth had significantly higher connectedness to school ($\beta = .523$, $p = .001$) and connectedness to teachers ($\beta = .356$; $p =$

.059). And there was a trend toward higher connectedness to the future (β = .183, *p* = .107) than the untreated controls. The finding of connectedness to teachers in these analyses is consistent with the separate cohort analyses in Table 4.1. Improvements in connectedness to school and to the future are not. And, unlike in Table 4.1, there were no effects of program participation on connectedness to friends or parents.

Table 4.2
ANOVA's for Effect size using Pooled data from MI

UTSA Variables: ANOVA Entry to Discharge	Post-YAP Mean	Post SD	Pre-YAP Mean	Pre SD	F	*p*	*d*
Connectedness to Friends	20.92	5.18	20.14	5.10	2.09	0.15	**0.34**
Connectedness to Family	22.00	5.10	22.08	4.68	0.08	0.77	-0.04
Connectedness to School	20.65	5.12	20.42	5.02	0.31	0.85	0.10
Connectedness to Teachers	21.24	5.22	20.74	4.79	1.40	0.24	**0.22**
Connectedness to Future	22.44	4.38	22.67	3.83	0.48	0.49	-0.11
Misconduct	4.65	3.94	5.59	4.17	**4.72**	**0.03**	**-0.47**
Ed Expect Finish HS	1.19	0.29	1.22	0.28	0.47	0.50	-0.04
Ed Expect Start college	2.90	0.95	2.86	1.05	0.26	0.61	0.04
Ed Expect Finish College	1.67	0.33	1.65	0.35	0.53	0.47	0.04
Work Plan Find a Job	3.39	0.76	3.41	0.74	0.23	0.63	-0.03
Expect Success at Work	3.46	0.66	3.56	0.65	1.94	0.16	-0.12

Table 4.3
ANCOVA for UTSA Variables (Controlling for Age, City, Sex, Race)

UTSA Variable (pooled imputation samples) from Entry to Discharge	Post-YAP Mean	Post SD	Pre-YAP Mean	Pre SD	F	*p*	*d*
Connectedness to Friends	20.92	5.18	20.14	5.10	1.56	0.13	.14
Connectedness to Family	22.00	5.10	22.08	4.68	**2.38**	**0.01**	**.17**
Connectedness to School	20.65	5.12	20.42	5.02	0.86	0.58	.10
Connectedness to Teachers	21.24	5.22	20.74	4.79	0.98	0.46	.11
Connectedness to Future	22.44	4.38	22.67	3.83	1.30	0.24	.13
Misconduct	4.65	3.94	5.59	4.17	1.77	**0.09**	**.15**
Ed Expect Finish HS	1.19	0.29	1.22	0.28	0.79	0.64	.10
Ed Expect Start college	2.90	0.95	2.86	1.05	1.69	**0.07**	**.14**
Ed Expect Finish College	1.67	0.33	1.65	0.35	1.03	0.43	.11
Plan Find a Job	3.39	0.76	3.41	0.74	0.97	0.48	.11
Plan Success at Work	3.46	0.66	3.56	0.65	1.24	0.30	.12

Overall, the strongest evidence of program-related improvements was in connectedness to teachers. There was no reliable evidence of negative effects of program participation. And although there was evidence of benefits of program participation on the other four outcomes, it was not consistent across methods. Evidence from the hierarchical analyses suggests participation in YAP was associated with increases in connectedness to the future and to school, but not in the separate cohort comparisons. The hierarchical analyses also provided no corroborating evidence of the change in connectedness to friends and family revealed in the separate cohort comparisons.

Results for expectations and misconduct. Reported below are cross-cohort and pooled cohort change estimates on four measures of conventional activity and one measure of unconventional, deviancy-related outcome. Table 4.4 provides the effect sizes for each of these pre (entry) to post (discharge) estimates. Descriptive statistics for these tests are in Table 4.2 and 4.3. As before, the top five rows provide change estimates for the five cohort groups separately. Pre-post (within-person; PP) change and cross-cohort treatment effects (TXCT) are presented as standardized mean difference effect sizes (d). The bottom three rows list effect sizes for pooled cohort pre-post comparisons PP t-tests and ANOVA/ANCOVA in which missing data were imputed

Expects to finish high school. The pre-post t-tests and cross-cohort effect sizes show evidence for change, but in inconsistent directions across cohorts. The ANOVA and ANCOVA do not show change.

Expects to start college. The individual pre-post t-tests effect sizes show evidence for change within cohorts, but the within-youth (pre-post) differences are in inconsistent directions and so the pooled effect size is negligible. Although the ANOVA and ANCOVA do not show change, the average cross-cohort estimate of between group differences is positive (d = .17).

Expects to finish college. The individual pre-post t-tests effect sizes show evidence for change within cohorts, but the within-person (pre-post) differences are in inconsistent directions and so the pooled effect size is negligible. Although the ANOVA and ANCOVA do not show change, the average cross-cohort estimate of change is positive (d = .24), as it was for expecting to start college.

Expects to find a job. The pre-post and cross-cohort effect sizes reveal evidence of change but in inconsistent directions. ANOVA and ANCOVA show no change.

Table 4.4
Cohen's d effect sizes for Pre-Post Change T-Tests and Untreated (Pre-) v. Treated (Post-scores of Prior Cohort) Difference ANOVAs, ANCOVAs, across Cohorts

	Finish High School**		Start College		Finish College		Find a Job		Misconduct*	
Cohort Number	PP	TX CT	PP	TX CT	PP	TX CT	PP	TX CT	PP	TX CT
C1 n = 37	.07	.06	**-.28**	.03	**-.31**	**.21**	**-.38**	-.06	-.04	**-.40**
C2 n = 40	-.07	**-.22**	**.26**	**.52**	**.42**	**.61**	-.07	-.04	**-.53**	**-.23**
C3 n = 34	**-.55**	**-.30**	**.57**	-.08	**.72**	.01	**.28**	**.33**	**-.51**	**-.43**
C4 n = 26	**.20**	**.16**	**-.18**	**.22**	-.19	.11	.00	-.05	-.02	**.25**
C5 n = 27	**.28**		-.07		-.19		**.15**		**-.14**	
ANOVA *d*		-.04		.04		.04		-.03		**-.47**
ANCOVA *d*		.10		.14		.11		.11		-.15
Pooled PP/TC *d*	-.02	-.08	.06	**.17**	.09	**.24**	.00	.04	**-.25**	**-.20**

Notes: PP denotes the effect size for a pre-post t-test comparing the average difference across youths' pre- to post scores within a single cohort. TXCT denotes effect sizes that reflect the magnitude of the difference between 4-month assessments of the treatment group and pre-treatment assessment scores for the subsequent cohort of youth (untreated comparison group). Positive scores indicate increases or higher scores for the treated youth. Below these are pooled sample untreated to treated group differences estimated using ANOVA and using ANCOVA. The bottom row provides the average pre-post difference across all individuals and under TXCT is the average effect size across the four treatment comparisons. All reflect Cohen's d as the effect size, for which d = .20 is "small," and a d= .40 to .50, is "moderate" in size. For the top five rows, bolded scores signify non-negligible (small or d $\geq$ .20) differences. In the bottom three rows, bold reflects statistical significance of $p \leq .05$. ** Values are square root inversion of original scores created to approximate a normal distribution. Negative scores indicate an increase in expectations of finishing high school among treated youth.

Misconduct. All indicators suggest declines in self-reported misconduct are associated with program participation. Only three of five pre- post comparisons met the d = .20 criterion, but this is a phenomenon that generally increases over time. Four of five cohorts showed small (d = -.14) to moderate (d = .53) cross-cohort treatment effect differences on misconduct. The average change from entry to discharge on misconduct also was meaningful (d = -.25), as was the averaged between-groups/cohort difference (d = -.20). The ANOVA for between group differences (d = -.47) was statistically significant, and at the p < .10 criterion the ANCOVA also suggest small decreases (d = -.45) in misconduct were associated with program participation once the contributions of sex, age, city and race are accounted for (d = .15). However because the effect size for the ANCOVA controlling for age was a third that of the ANOVA, we felt tests to rule out

maturation as a viable validity threat were necessary. These are conducted in the following section. Overall, however, the evidence for change on this outcome is strong.

Hierarchical analyses. The hierarchical models, which nested, treated and comparison youth data pooling all youth into one test did not yield significant main effects of treatment for any of these five variables. Only on Misconduct was there even a trend. For self-reported misconduct the change coefficient was -.126, suggesting lower misconduct among treated youth at discharge, but the significance level only reached a *p* = .199 trend.

In summary, four-month changes in expectations regarding academic and vocational success were generally in consistent and unconvincing. Only expectations of going to college and graduating from college differing in two cohorts each, but had the mean effect size averaged across cohorts being .17 and .24 respectively, with scores being higher for treated youth. However, the analyses of program-related change on misconduct were compellingly consistent, with only the ANCOVA and multilevel analyses suggesting smaller effects than other estimates, which may reflect the role of starting risk may be a rival interpretation of the findings. Note, age effects on these changes were not apparent in earlier maturation tests, so this validity threat does not seem to be at play. Given the role of initial rates of misconduct on declines we are associating with program participation, the possibility of regression to the mean needed to be addressed.

Rival internal validity check for regression to the mean on misconduct. One of the main critiques of a RIC design, particularly when used to assess a program's impact on an outcome variable (like "misconduct"), which is also the primary criterion for admission to the program is regression to the mean effects. People who enter a program reporting extremely high (or low) levels of a problematic behavior or state (particularly one targeted for reduction by program engagement), will over time naturally move toward a less extreme position. (This is well evidenced in the psychotherapy literature, though the therapy literature typically is based on the change observed among those who voluntarily seek therapy not simply score at the extremes on some phenomenon like depression or anxiety.) Furthermore, individuals whose scores are most extreme may reflect inaccurately high scores that also will regress to their true score, which is closer to the group mean. If this is the case, then you would expect those individuals who are so heavily engaged in misconduct as to be sent to YAP would naturally decline in misconduct—how much higher could they go? This could appear like a treatment effect.

Therefore, the positive misconduct results reported above are subject to the alternative explanation of regression to the mean. To address this problem, we first looked at the starting scores on that variable for individuals entering the program. This

revealed whose reports were most extreme and subject to regression effects. We separated the youth who were in the middle, those one SD above and those one SD below the mean into three groups reflecting extremely high, average, and extremely low scores. Then we created a variable of change over time on misconduct by subtracting the post score (discharge) from the pre score (entry). (This required us to impute the missing data for those who did not complete the program, many of whom could have been initially extreme responders. We then used the 5 imputed datasets to conduct our comparisons.) Then we examined their change over time both visually and statistically using ANOVA to compare scores across groups. Table 2.26 provides means for teach group across the five datasets.

Table 4.5
Mean Misconduct at Pre-test for High, Average, and Low Scoring Groups

Dependent Variable: Pre Survey Misconduct_No Outliers

Imputation Number	Misconduct reg to mean test grps	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Original data	low misconduct	1.081	.312	.465	1.697
	mid misconduct	5.012	.208	4.601	5.423
	hi misconduct	12.014	.330	11.362	12.666
1	low misconduct	1.077	.304	.477	1.677
	mid misconduct	5.057	.202	4.657	5.456
	hi misconduct	11.985	.312	11.369	12.601
2	low misconduct	1.053	.300	.461	1.644
	mid misconduct	5.044	.195	4.660	5.429
	hi misconduct	12.068	.308	11.461	12.676
3	low misconduct	1.026	.299	.434	1.617
	mid misconduct	5.045	.198	4.654	5.436
	hi misconduct	12.096	.312	11.481	12.711
4	low misconduct	1.105	.304	.506	1.705
	mid misconduct	5.000	.196	4.613	5.387
	hi misconduct	11.927	.316	11.303	12.552
5	low misconduct	1.077	.301	.483	1.670
	mid misconduct	5.100	.198	4.709	5.491
	hi misconduct	11.985	.317	11.358	12.611
Pooled	low misconduct	1.067	.303	.473	1.662
	mid misconduct	5.049	.202	4.654	5.445
	hi misconduct	12.012	.322	11.381	12.644

As expected, those who entered the program with the highest scores on misconduct showed the greatest declines (see Figure 4.1 below). Both the highest and lowest scoring groups regressed to the mean, with the lowest scoring group reporting an increase. This suggests that the statistical issue of regression to the mean (due to extreme

responding at pre) does seem to have influenced both extremely high and low scoring youth to regress toward the mean. However, the slope or magnitude of change for those reporting the most misconduct at pre is twice that of the low misconduct group. Most important, evidence that observed change is not a statistical artifact of regression to the mean is seen in the reductions in misconduct for the group whose scores were average at entry. This suggests that at least some of the declines in misconduct among those initially reporting the most can be due to program participation.

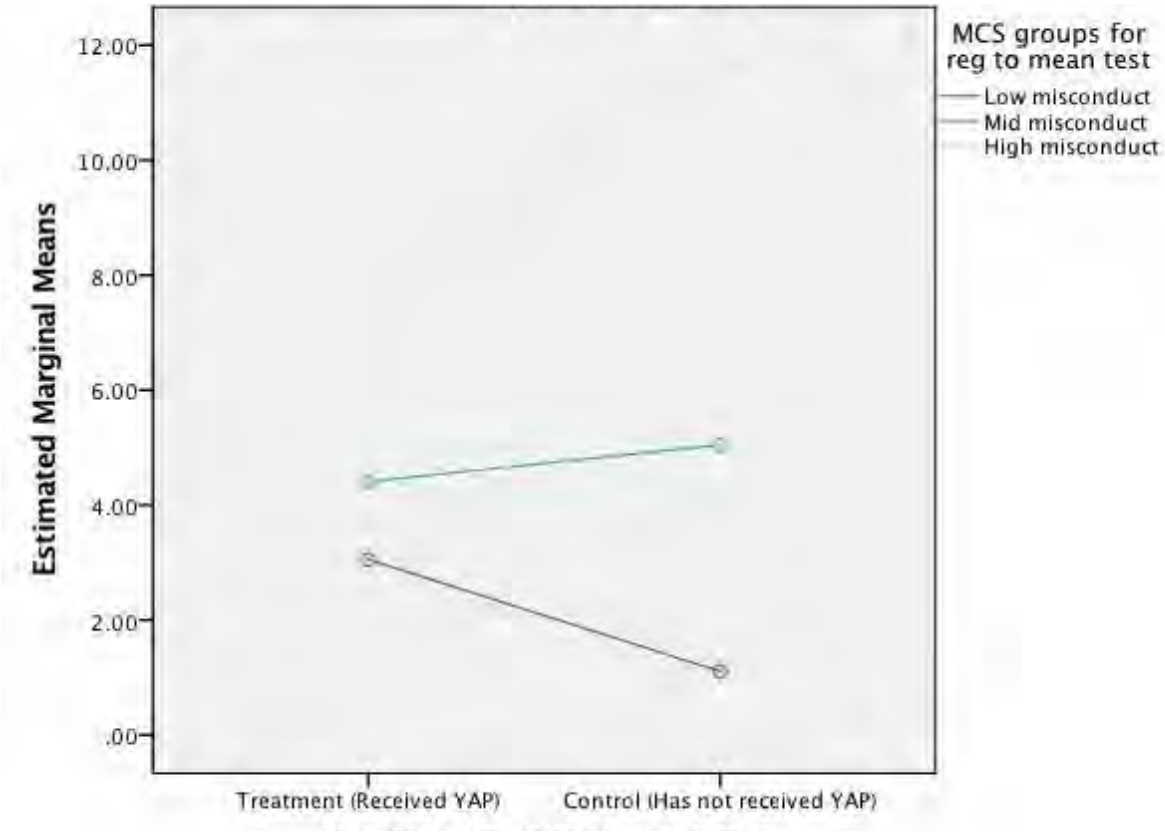


Figure 4.1. Within-person (Pre- to Post-test) Change for Youth Starting at High, Average, and Low Levels of Misconduct at Pre-test (Entry).

To effectively rule out regression to the mean processes, analyses looking at just those youth in the middle group were done (n = 174; 87 per treatment group when estimates for those in the average-at-entry across imputation datasets were pooled) to estimate the changes between treated and untreated groups of youth with similar, average entering misconduct scores using a comparably sized sample. Using the five multiply imputed datasets, and analyzing data separately by cohort to remove dependency from the main effect tests of treatment, a one-way ANOVA was conducted.

Table 4.6
Means for Treated and Untreated Youth on Misconduct

Dependent Variable: Misconduct

Imputation Number	Control (0) and Treatment (1)	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Original data	Control (Pre)	4.980	.235	4.515	5.445
	Treatment (Post)	3.723	.342	3.045	4.400
1	Control (Pre)	4.923	.290	4.350	5.495
	Treatment (Post)	4.467	.290	3.895	5.040
2	Control (Pre)	5.003	.261	4.487	5.518
	Treatment (Post)	4.468	.261	3.953	4.983
3	Control (Pre)	5.007	.168	4.676	5.338
	Treatment (Post)	4.743	.203	4.341	5.145
4	Control (Pre)	5.039	.248	4.549	5.528
	Treatment (Post)	3.868	.248	3.379	4.358
5	Control (Pre)	4.907	.284	4.346	5.468
	Treatment (Post)	4.516	.284	3.955	5.077
Pooled	Control (Pre)	4.976	.262	4.462	5.489
	Treatment (Post)	4.413	.440	3.422	5.403

The pooled mean misconduct scores across all five cohorts and all five imputed data sets yielded means of 4.98 for the untreated comparison group and 4.41 for the treated group. The pooled effect size across these comparisons was an eta-squared of .023 before separating groups into cohorts, and .054 when effect sizes were estimated within each group (independent data) and then averaged. These translate to small to moderate size effects of program participation in decreasing misconduct. This effect size reflects change illustrated by the green line in Figure 4.1 above. The magnitude of this effect matches that observed in the earlier estimates of pre-post and across-cohort comparisons (see table 4.4).

Controlling for the influence of sex and age on the averaged differences within cohorts between treatment and controls yielded estimated means of 5.02 for the untreated comparison group and 4.36 for the treated group, which is larger than the difference without these covariates. This is not consistent with the ANCOVA analyses for the sample as a whole reported above using the d-statistic for effect size (d = -.14), but those analyses did not impute missing values. Arguably, these are more reliable because those covariates were balanced when the cohorts were pooled (because the same kids in pre and post, so only age differed but it did so the same for all youth). More importantly these analyses do not include the low and high scoring individuals, whose opposing change coefficients simply add error to the model reported earlier.

These analyses provide good support for the argument that program effects should not be attributed to regression to the mean processes.

Changes in crime, education and job activity measures. These tests represent change from entry to discharge. Effect sizes are presented in the following table and summarized briefly below.

Living situation. The t-tests revealed inconsistent effect sizes across cohorts, negligible change, and in different directions, particularly for the treatment-untreated differences. The average effect sizes for both t-tests and the ANOVA suggest no change over time nor between treated and untreated youth.

Education situation. Results of the t-test and ANOVA suggest positive changes in education situation, indicating that participating in YAP increased the youth's standing in school participation from not enrolled towards full time attendance, graduation, or college enrollment.

Employment status. Results of the t-test and ANOVA suggest positive changes in employment status as a function of participating in YAP. These scores would be consistent with YAP participation increasing the youth's interest and attempts to find employment.

Table 4.7
Cohen's d effect sizes for Pre Post and Treated vs. Untreated Difference T-Tests for RIC Design, with YAP-Collected Data: At Entry (Untreated) vs. Discharge (Treatment)

	Living Situation		Education Situation		Employment Status		Serious Disposition	
	PP	TX CT	PP	TX CT	PP	TX CT	PP	TX CT
C1 n = 37	.09	**.49**	**.61**	**.51**	**.86**	**.60**	**-3.04**	**-3.58**
C2 n = 40	-.02	-.12	**.47**	-.03	**.31**	.09	**-3.09**	**-3.57**
C3 n = 34	.09	.07	**.31**	**.62**	**.71**	**.37**	**-2.17**	**-2.29**
C4 n = 26	**-.27**	**-.69**	**1.01**	**.65**	**.34**	**.97**	**-1.86**	**-1.07**
C5 n = 27	**-.42**		**.45**		**1.00**		**-1.46**	
ANOVA		-.001		**.51**		**.59**		**-2.45**
Pooled PP/TC *d*	-.11	-.06	**.57**	**.44**	**.65**	**.51**	**-2.32**	**-2.63**

Notes: PP denotes the effect size for a pre-post t-test comparing the average difference across youths' pre- to post scores within a single cohort. TXCT denotes effect sizes that reflect the magnitude of the difference between 4-month assessments of the treatment group and pre-treatment assessment scores for the subsequent cohort of youth (untreated comparison group). Positive scores indicate increases or higher scores for the treated youth. Below these are pooled sample untreated to treated group differences estimated using ANOVA and using ANCOVA. The bottom row provides the average pre-post difference across all individuals and under TXCT is the average effect size across the four treatment comparisons. All reflect Cohen's d as the effect size, for which d = .20 is "small," and a d= .40 to .50, is "moderate" in size. For the top five rows, bolded scores signify non-negligible (small or d $\geq$ .20) differences. In the bottom three rows, bold reflects statistical significance of $p \leq .05$.

Serious disposition. Results for both within person change across treatment (PP) and treated-untreated differences were large and consistently in the same direction. These findings suggest that participating in YAP was associated with a statistically significant and large decrease in the most serious disposition (status offense, misdemeanor, felony) on the youth's record relative to the untreated comparison youth.

Table 4.8
ANOVA for YAP Variables Entry to Discharge

YAP Variables Entry to Discharge (DC)	DC	DC SD	Pre-(Entry)	Pre SD	F	*p*	*d*
Living Situation	4.51	1.10	4.51	1.18	.05	.86	-.001
Serious Disposition	.32	0.75	2.23	0.80	491.70	.00	**-2.45**
Education Situation	2.92	0.65	2.55	0.78	21.27	.00	**.51**
Employment Status	1.57	0.83	1.15	0.58	28.72	.00	**.59**

Changes from Entry to 12 Months Post Discharge in Education, Employment and Job Activity

The analyses in the table below represent change during the period from entry into YAP to 12 months post-discharge. Only data collected by YAP is available for the one-year follow up analyses. These tests do not control for length of treatment.

Living situation. Regarding their living situation, the results are inconsistent in regards to change, suggesting youth were not changing in their living situations between entry to post-discharge periods. However, there was constrained variance in their living situations data.

Education situation. Results of the t-test and ANOVA suggest positive changes in education situation from entry to post-discharge were related to participation in YAP. Program participation predicted increases in school participation from not enrolled (low scores) towards full time attendance, graduation, or college enrollment.

Employment Status. Results of the t-test and ANOVA suggest positive changes in employment status from entry to post-discharge were observed and were associated with participation in YAP. These scores suggest YAP graduates reported greater interest in and more attempts to find employment after treatment than youth who had not completed YAP.

Table 4.9
Cohen's d effect sizes for Pre Post and Treated vs. Untreated Difference T-Tests for RIC Design, with YAP-Collected Data: At Entry (Untreated) vs. 12 Months Post-Discharge (Treatment)

	Living Situation		Education Situation		Employment Status		Serious Disposition	
	PP	TX CT	PP	TX CT	PP	TX CT	PP	TX CT
C1 n = 37	.01	**.40**	**.60**	**.74**	**1.65**	**1.64**	**-1.83**	**-2.07**
C2 n = 40	-.13	-.15	**.45**	**.32**	**.88**	**.99**	**-1.64**	**-1.97**
C3 n = 34	**.17**	**.19**	**.19**	**.62**	**.80**	**1.02**	**-2.34**	**-1.99**
C4 n = 26	-.15	**-.53**	**.69**	**.61**	**.93**	**1.71**	**-1.29**	**-.66**
C5 n = 27	**-.55**		.05		**1.33**		**-1.00**	
ANOVA *d*		.02		**.56**		**1.36**		**-1.64**
Pooled PP/TC *d*	-.13	-.11	**.53**	**.57**	**1.12**	**1.34**	**-1.62**	**-1.67**

Note: PP notes the effect size for a pre-post t-test comparing either a single cohort's pre-survey scores to the same cohort's 4-month survey scores, or (on bottom row) for the average pre- to post-assessment effect size across the five cohorts. TXCT effect sizes reflect the magnitude of the difference between 4-month assessments of the treatment group and pre-treatment assessment scores for the subsequent cohort of youth (untreated comparison group). These use Cohen's *d* as the effect size, for which *d* = .20 is "small," a *d*= .40 to .50, is "moderate" in size.

Serious disposition. These findings suggest that from entry to 12 months post-discharge, the severity of a youth's disposition (status offense, misdemeanor, felony) decreased as a function of participating in YAP.

Table 4.10
ANOVA for YAP Variables Entry to 12-Months Post-Discharge

YAP Variable Entry to Post-DC	Post-DC	Post SD	Pre-YAP	Pre SD	F	*p*	*d*
Living Situation	4.53	0.91	4.51	1.18	1.10	.48	.02
Serious Disposition	0.77	0.97	2.23	0.80	222.55	.00	**-1.64**
Education Situation	3.07	1.07	2.55	0.78	026.37	.00	**.56**
Employment Status	2.30	1.04	1.15	0.58	155.43	.00	**1.36**

Hierarchical models to address dependencies in the data. Dyadic data analyses were used to address the issue of statistical dependency in the data across the treatment and untreated comparison group. These analyses were conducted on two 12 month follow-up variables, (1) serious crime and disposition and (2) educational engagement.

For each variable, multilevel analyses allowed shared variance across treatment and untreated comparison groups to be apportioned into random error, thereby freeing tests of treatment effects from statistical dependency. In addition, by controlling for age of assessment a further separation of the effects of age on observed differences from the estimates of differences between the treatment and untreated comparison group was achieved. Separate analyses excluded individuals whose ages fell outside the range of the complementary treatment group's age (e.g., those treated youth four months older than the oldest comparison youth).

Serious disposition. First a simple two-level model predicting variability in rates of serious disposition for treatment and untreated comparison group members was run without any controls for age. The raw data included 247 individuals from 155 clusters.

The degree of dependency in the data was relatively small. The intraclass correlation (or amount of variability in serious disposition that was due to the clustering of treatment and comparison observations within individuals, random effects) was .012. These treatment-related effects are estimations after the bias in standard errors due to data dependency was removed through the modeling of the dependencies in the data. The model fit indices were positive suggesting this basic model was a good fit with the data. The CFI of .99 and significant Chi-square test coefficient (55.79) both suggested a good model fit.

The unstandardized slope for treatment, which reflects the unadjusted, unstandardized mean difference between comparison and treatment youth indicated that one year post-treatment YAP graduates were -1.762 (out of a range of 4 points) lower in their scores on most serious disposition than the untreated youth. The amount of variability in degree of serious disposition among youth that was associated with program participation was 26%, which is a statistically significant and large effect.

The age of assessment was included in a subsequent run of the basic model described above. The results were similar and the declines in most serious disposition associated with participation in YAP were slightly greater than in the more restricted model above. The between group difference was -1.94 with a non-significant contribution of age in the prediction of most serious disposition. Age was significantly associated with treatment condition ($\beta = .49$). Therefore the fixed effect analyses (with missing data imputed) in the prior section are consistent with these multilevel analyses. Both of which suggest that age the association between age and disposition is negative ($r = -.72$) for the sample as a whole; this, however, is a function of the association between treatment groups and age (with treatment group members being on average a year and four months older than untreated youth in the untreated comparison group). But given that rates of engagement in criminal activity and court involvement increase normatively, the treatment effect appears to have suppressed this normative increase.

Educational engagement. Multilevel analyses of the association between engagement in education in terms of attendance also were conducted in which treatment group conditions were nested within individuals. Analyses were run on for those with scores only the first three categories—no attendance, part-time attendance, and regular attendance in school—which are obviously ordinal were used. But the category, "Earned GED," is not. It cannot be effectively argued that earning a GED reflects greater educational engagement than attending school regularly. Similarly, having graduated from high school—the fifth category—is not an option for younger youth from whom this entry would be censored. Therefore, in this model, educational engagement is actually a measure of attendance.

One other issue that arises when comparing the treatment conditions on educational engagement, and in particular across age-cohorts, is that just as there were more treatment individuals who were old enough to have graduated, there were not enough of the treated individuals (whose educational engagement was assessed one year post-discharge) below the age of 14 and a half to make appropriate comparisons with the untreated comparison group whose lower age range was 13. Therefore, the analyses on educational attainment were run only for two of the six age cohorts of research participants. The cohort of fourteen to fifteen and a half year old participants and the cohort of fifteen and a half year olds to seventeen year olds were included in these analyses.

A simple two-level model predicting variability in rates of educational engagement (namely attendance) for treatment and untreated comparison group members was run that included the youths' ages at the time of assessment of educational engagement. The raw data included 164 individuals from 124 youth-Advocate clusters or pairs. The degree of dependency in the data was larger for educational engagement than it was for serious disposition. The intraclass correlation or amount of variability in educational engagement that was due to the clustering of treatment and comparison observations within individuals (random effects) was .056. This means it was indeed important to capture this dependency through the multilevel analyses described below.

Most model fit indices were positive suggesting this treatment effect model was a good fit with the data. These indices include a CFI of 1, TLI of .99, RMSEA = .03, and significant Chi-square test coefficient (437).

The unstandardized slope for treatment, which reflects the unadjusted mean difference between comparison and treatment youth, controlling for variability in age, indicated that one year post-treatment YAP graduates were 1.08 (out of a range of 3 points) higher in their scores on educational engagement than those untreated youth which. This is a large effect.

In summary, changes in engagement in actual behaviors related to school and employment associated with program participation were consistently large and positive across cohort and assessment methods. Pre-post, between group and multilevel analyses

revealed strong evidence of improvements in educational engagement (attendance) and job seeking (or job securing) following participation in the YAP program. The long-term changes observed sixteen months post-entry on educational and vocational engagement replicated those at four months with change observed consistently across all methods of analysis suggest reliable program-related benefits on engagement in school and employment.

Conclusions from Study I

These RIC analyses suggest positive effects of youths' participation in YAP on several outcome variables, including the prosocial behaviors/attitudes (connectedness), self-reported misconduct, and educational, work, and crime involvement after four months of program participation. Comparisons between treated and untreated youth on employment and educational engagement as well as most serious disposition also were found at 12-months post-discharge.

Prosocial behavior/connectedness to teachers. Regarding short-term changes on prosocial behavior (connectedness to friends, family, school, teachers, and the future) associated with program participation, statistically significant changes or small or larger between-group effect size estimates of differences were observed for the whole sample on four of these five measures using analyses of covariance (between-group differences) or t-tests of effect sizes (for pre-post change) (see Table 2.24). These results reveal between-group differences corroborated pre-post increases in connectedness to friends and connectedness to teachers that were small to moderate in size. Group differences in connectedness to family were revealed in an analysis of variance, but this difference between treated and untreated groups was no longer statistically significant once age, gender, and starting levels of misconduct were accounted for.

Of these differences, however, only increases in connectedness to teachers was corroborated by between-group differences estimates using hierarchical analyses that accounted for the statistical dependency caused by including youth's pre and post in the same analyses of between-group differences (by nesting assessments within individuals). Hierarchical analyses revealed that program participation predicted significantly higher reports of school connectedness and connectedness to the future among program completers than untreated youth, but these findings were not evidence in the simpler pre-post and between group (cross-cohort) comparisons. Conversely, no effects of program participation on connectedness to family or friends were revealed in the multilevel analyses despite being found in the simpler pre-post and between group analyzes. These mixed results render evidence of change equivocal.

Educational, vocational, and court involvement. Using the data that is regularly collected by YAP and that is more objective in nature, we found more consistent evidence of change across analysis methods. Looking at youth's living situation, educational situation/engagement (attendance), employment status (interest in/actively

looking for), and serious disposition with tests of both within person (pre-post) change and of differences between treatment/untreated comparison groups.

On all of these, except for living situation, we observed improvements while in YAP (individual-level pre-post change) and evidence of between group differences that were related to program participation, both favoring those who participated in YAP. In addition, the benefits of program participation on these three outcomes was both large and nearly identical across the two approaches to estimating change.

Finally we conducted 12-month follow-up tests of change associated with program participation on the same, regularly collected YAP outcomes. These tests provided confirmation of the maintenance of several positive outcomes associated with program participation. Benefits of YAP participation on YAP graduates' educational situation at discharge held constant at one year post-discharge. However the benefits on educational status associated with program participation suggest YAP graduates improved considerably during the one year after completing the program.

Although the benefits of program participation on YAP graduate's self-reported status offenses, misdemeanors, and felonies ("serious dispositions") one year-post discharge were not as large as those at the end of program participation (at discharge), these long-term benefits of program participation are still quite large. Consistent with estimates of individual changes from program entry to discharge, the results revealed a larger decrease in self-reported status offenses, misdemeanors, and felonies ("serious dispositions") from entry to twelve months post-discharge. Analyses also revealed moderate increases in both the youth's educational situation and in their interest in (or effort toward) finding employment one-year post program completion.

Therefore, the long-term, 16 months post entry, gains in educational engagement and relative reductions in criminal activity corroborated effects found at four months. Large benefits of program participation on YAP graduates' self-reported criminal disposition (rates of status offenses, misdemeanors, and felonies) were observed both at program completion and one year-post discharge.

Misconduct. The short term, entry-to-discharge changes associated with participation in YAP on misconduct represent statistically significant differences between the untreated comparison group (pre-treatment) and the treatment group (age-matched program completers) at discharge that suggest positive effects of program participation on declines in self-reported misconduct (see Table 4.4). Corroboration of change estimates was revealed in statistically comparable effects for differences between the treated and untreated comparison groups and for pre-post change within individuals. That is, on self-reported misconduct, the size of the average person's decline from entry to discharge on misconduct was similar to (or corroborated by) the magnitude of the differences between program participants and the counterfactual untreated comparison group.

It should be noted that these relationships between program participation and declines in misconduct are of a magnitude that is as large or larger than the effects of participating in comparable recidivism reduction mentoring programs (i.e., d = .20, or "small," to d= .40 to .50, or "moderate" effect sizes"), which tend to be more moderate in size,[7] and for school-based mentoring programs.[8]

Other than the program-related differences in academic and job-related behaviors from entry to discharge, the most convincing of the pre-post estimates of program-related change was in declines in self-reported misconduct. There was corroboration in the direction and magnitude of effect sizes (consistency) across pre-post t-tests, between group difference effect sizes for cohorts separately, between group average difference effect sizes two-way analyses of variance, and a complementary statistical trend found in the hierarchical comparisons. In addition, tests of maturation and regression to the mean could be inferred as ruling out these rival hypotheses of these changes associated with program participation.

Overall, the evidence of positive program effects, using the variables we included and the design we employed, are strongly positive and consistent. Some studies measure changes over time, as did this study. Other designs compare those treated to those non-treated, and we did this too. In combination, and with a number of rival hypotheses like maturation and regression to the mean ruled out, the use of the recurrent institutional cycle design yielded consistent evidence of positive effects of participation in YAP on the most relevant outcomes and across multiple statistical tests.

---

[7] Tolan P, Henry D, Schoeny M, Bass A (2008). Mentoring interventions to affect juvenile delinquency and associated problems. *Campbell Systematic Reviews*, 16. DOI: 10.4073/csr.2008.16

[8] Herrera, C., Grossman, J. B., Kauh, T. J., Feldman, A. F., McMaken, J., & Jucovy, L. Z. (2007). *Big Brothers Big Sisters school-based mentoring impact study*. Philadelphia: Public/ Private Ventures.

V. Study 2: Theory-based Program Practices Process Analyses

The YAP organizational mission is to provide court-involved youth with opportunities to develop personally, contribute to their communities, and have safe, effective, and efficient alternatives to institutional placement. Beginning at referral, Advocates focus on developing a relationship, strengthening family and community relations, and ensuring that youth address community service and court mandates. Adult Advocates are selected in terms of their similarity to the youth they serve (e.g., are recruited from the same zip-code as youth) to help ensure that service providers have a grassroots understanding of community assets, challenges, and opportunities.

The primary goal of the second study was to assess whether what the youth and their mentors do together explains program-related changes and whether pre-match characteristics of youth or their mentors predict the kinds of activities they did together. A series of analyses are presented that provide insights into who does what and why some kinds of activities, at specific points in time during the match, may be most useful.

Having found in Study 1 reliable associations between program participation on declines in misconduct, this variable was selected as the primary criterion variable for examining the role of participant characteristics and activity choices in program outcomes. Structural equation modeling analyses were conducted for the three separate sets of analyses in Study 2 in order to assess whether the relationships between activities and changes in misconduct were moderated by youth or Advocate characteristics, particularly youth's starting risk level and Advocates' prior experience as educators and level of their educational attainment.

These tests are guided by a theoretical model that presents several activity types, reflecting three dimensions of mentoring, that have been described in the mentoring literature as most useful or important in influencing mentoring relationship quality and mentoring program outcomes. This theory, called the TEAM framework, is described in the next section. Based on the nature of advocacy, the YAP program model, and the youth clients served by YAP, the analyses in this chapter focus on three main types of activities—play, problem-focused discussions, and casual conversations with youth.

Following the presentation of the TEAM framework and a description of what the framework says about the role of when specific activities happen during the relationship, three hypotheses are presented. These test the value of play and problem-focused interactions, the importance of shifts in activities across the life of a match, and the role of youth and adult characteristics. These are tested using simple structural equation model analyses which are conducted twice, once using observed scores and once using factor scores. The series of analyses with each of these two types of activity variables starts with just activities predicting misconduct, then adds in youth characteristics, and then includes Advocate characteristics to test for direct and indirect effects of Advocates' level education and prior teaching experience.

The two sets of structural path models are preceded by a simple test of growth over time in misconduct by examining the slopes and intercepts of misconduct at entry, midway through the match, and at discharge. Then, correlational analyses are presented that convey the associations between participant characteristics and starting levels of misconduct. These analyses include only youth and Advocate sex and age and both Advocate education and teaching experience. Youth sex and both Advocate educational characteristics were correlated and subsequently found to be valuable exogenous predictors in the models explaining variability in rates of misconduct at discharge.

There are two parallel sets of path models run, both of which test the same sequence of main effects of activities, the contribution of youth characteristics, and then the indirect and direct effects of Advocate education variables. The difference between the two, parallel analyses is that the first set of analyses conducted used observed activity frequency scores that averaged the frequency of all indicators of each activity type to get a mean score.

The second set of analyses were conducted to corroborate the observed score findings using factor scores to measure activities. Using factor scores allowed each activity indicator (see Table 5.2) to make a unique contribution to that activity factor score, rather than given each one the same weight (even though matches likely rarely did all of them). These factors score were estimated separately for Advocates and youth and separately at the two time points. This provided useful information about how youth differed in their views of what happened at the two points in time.

The mentoring literature has long suggested that mentoring effectiveness varies as a function of specific shifts in the focus of conversations and activities across the first and second half of the relationship.[9] Therefore, the next step was to try to better understand whether changes in types of interactions across the treatment period could explain how misconduct may be influenced by the mentoring and advocacy relationship interactions.

The framework that summarizes prior research on the value of changes in the rates of different types of activities across the life of a match in effecting program outcomes is presented next. This framework, which was created to account for the dimensions of the activity focus, purpose, and collaboration in mentoring programs' success, was used to generate three hypotheses that guided the analyses in Study 2.

[9] Karcher, M. J. & Hansen, K. (2014). Mentoring activities and interactions. In D. L. DuBois, & M. J. Karcher (Eds.), *Handbook of youth mentoring* (2nd Edition; pp. 63-72). Thousand Oaks, CA: Sage Publications

Measuring Activities Using the TEAM Framework

The TEAM Framework[10] was used to conceptually organize the different kinds of interactions that take place between Advocates and youth in the YAP program. There are three main characteristics the TEAM Framework uses to describe interactions. *Activity Focus* is a term used to reflect how relational vs. directive the mentoring interactions are (the focus of activities, discussions)(see Table 5.1). For example, some interactions are more about talking and learning about one another while other interactions are about doing something together, or trying to solve a problem faced by the youth.

A second characteristic is about the nature of decision-making in the relationship. The *Activity Choice* or *Authorship* dimension reflects how collaboratively the conversation or activity decisions are made? Who "authors" or drives their story in terms of influencing what they do together (regardless of the focus of what they do)? Mentoring literature suggests that collaboration is critical to relationship development and ownership in a relational process. Collaboration happens when both make unique contributions to deciding what to do and the two create a new goal, rather than either one dictating what they do.

Last, the *Activity Purpose* characterizes the intended outcomes of a conversation or activity. In terms of the literature on delinquency, specifically that of Jessor's Problem Behavior Theory (Jessor & Jessor, 1971), two main factors contributing to delinquency: conventionality and unconventionality. Conventional activities, values, and goals reflect the conventions or values of adult society. Conventional interactions and relationships, usually with adults, serve serious, future-oriented, adult-held goals. In contrast, unconventional activities, values, goals and interpersonal relationships reflect or are organized around the priorities of youth, which tend to reflect peer interactions, present-focused activities, and the pursuit of fun.

When there is imbalanced investment by youth in conventional and unconventional activities, values, and goals, such that when a youth puts greater personal investment in unconventional activities, values, and goals, he or she is at much greater risk for criminal activity. Jessor and Jessor (1971) found that youth with few or minimal conventional connections or daily activities often harbor disdain toward adults, dismiss the consequences of their actions on their future liberties and successes, and challenge or subvert rules and laws.

The efforts of adults to understand the youth's investment and to give it validity through joining in the youth's world often can trigger the youth's movement back toward greater balance of conventional and unconventional investments in return. Therefore it was hypothesized that the inclusion of playful activities in the mentoring relationship would lead to stronger youth-Advocate relationships

[10] Karcher, M.J. & Nakkula, M.J (2010). Youth mentoring with a balanced focus, a shared purpose, and collaborative interactions. In *New Directions for Youth Development* (126) [special issue "Play, talk, learn: Promising Practices in Youth Mentoring,"] Jossey-Bass.

Table 5.1
Theoretically Evolving Activities in Mentoring (TEAM) Framework (Karcher & Nakkula, 2010)

Purpose	**Unilateral Authorship: Mentor driven**	**Collaborative Authorship: We focus (collaboration)**	**Unilateral Authorship: Mentee driven**	**Purpose**
Serves conventional (adult) purpose	(Adult-centric)	**Focus:** Minimally goal-directed/structured and/or highly relational	(Youth-centric)	**Serves playful (or youthful) purpose**
Adult-led spontaneous (non-relational)	**1. Preacher**	**2. Peer (classmate)** (is like Keller & Pryce's *Acquaintance*)	**3. Joker**	Youth-led spontaneous (non-relational)
Adult-oriented preventive and developmental activities or discussions (relational focus)	**4. Counselor**	***5.* Developmental Mentor** (Morrow & Styles' term *Developmental Mentor)* is like Keller & Pryce*'s "Friend*")	**6. Playmate**	Youth-oriented preventive and developmental activities or discussions (relational focus)
Conventional Skill Development **Purpose** relevant to adult/societal goals, interests, or beliefs about what the mentee needs (Primarily goal-oriented focus)	**7. Teacher** (Keller & Pryce's "*Tutor"*)	**8. Instrumental Mentor** (Hamilton & Hamilton's term *Instrumental Mentor is like* Keller & Pryce's, "*Sage*"	**9. Teammate**	**Playful Skill Development** **Purpose** relevant to the youths' goals, interests, or emphasizes outcomes now (Primarily goal-oriented focus)
Remedial/ Intervention-oriented: Serves adults' goals (goal-oriented)	**10. Colonel** (Boss, Vice Principal)	**11. Journeyman** ("Master Craftsperson" and Apprentice)	**12. Coach**	Remedial/ Intervention-oriented: Serves youths' goals (goal-oriented)
Serves conventional (adult) purpose	(Adult-centric)	**Focus:** Highly structured and goal-directed (and/or minimally relational)	(Youth-centric)	**Serves playful (or youthful) purpose**
Purpose	**Unilateral Authorship: "Mentor focused"**	**Collaborative Authorship: We decide together what to focus our time on**	**Unilateral Authorship: "Mentee focused**	**Purpose**

Especially for teens, for whom reciprocal adult-youth relationships signal a degree of respect and acknowledgement of the youth's value to society, engaging in activities like play can serve as a sign of respect and interest in the youth that can result in increased interest by the youth in adult-valued, conventional activities, and, conversely predict lower engagement in anti-adult, criminal misconduct. Play and problem-focused discussions, therefore, serve a different purpose but are related in that a balance between them is critical just as it is between conventional and unconventional investments.

Using the TEAM Framework to Predict Change from Mentoring Activities

A mentoring activity log was used to track what happened in relationship over time. Using that log, mentors and mentees reported the frequency of play, talking, problem-solving, and advocacy activities at two and four months after matches began. Responses ranged from Never (1) to Very Often 5. This log is presented in Table 5.2.

One mentoring activity that has received empirical examination for its value in effecting change, especially with younger youth in community based mentoring programs, is play. In this study play was measured using the following activities: Playing sports, Athletic activity, or Outdoor game; Creative activities (do art, read for fun, write a story or song); Playing cards, board games, computer games; Going to a park, museum, movie, community/cultural event, or college. These are shown in the log in Table 5.2.

The developmental and prescriptive mentoring approaches. The first hypothesis tested was that the more misconduct a youth reported when entering the YAP program the more frequently the youth and mentor would focus on serious and future-oriented (conventional) interactions, particularly early in the relationship. Mentors might immediately focus on the youth's problems because they think their mentoring relationship should be about helping the youth change how they think, act, and plan in order to be more successful in society in the future.

Conversely, youth entering the program reporting having engaged in less misconduct might have Advocates who, at the start of the relationship, feel comfortable to engage in friendship building activities, such as by playing games and doing sports together, to build a solid friendship that could later incorporate more effective problem-solving. Mentors who engaged in this "first-play" and "then-work" approach would be taking the approach Morrow and Styles found most effective in their study of the "Big Brothers Big Sisters" program, which they called the *developmental mentoring style*. Therefore, knowing whether youths' risk-taking level was the primary determinant of what happened in the match was a critical first step in understanding the link between activities and program outcomes.

The approach Morrow and Styles found least effective with children in the Big Brothers Big Sisters program, and which is commonly taken by mentors, is to start off the relationship with a goal-directed or problem-focused focus, paying most attention to things that reflect adult-relevant problems and attempt to effect success in the future. This approach, however, usually starts by assessing the youths' deficits to identify what

problematic behaviors, attitudes or knowledge deficiencies need to be corrected. Taking this deficit view and imposing a problem-solving approach immediately can be off-putting to youth, and is what Morrow and Styles (1995) called a *prescriptive* approach.

The prescriptive approach may be even more commonly taken when mentoring youth involved with the courts. Those mentors feel compelled to explain to youth the problems that can result from their misbehavior and underachievement. This *correcting the deficits* approach may be invited by a youth who is seeking out this information and assistance from the mentor. But more often the decision to start off with this approach is not collaboratively made. But a conventional focus can be useful when it reflects a collaborative goal-directed focus like that of a master-and-apprentice type relationship, which is shown in the center column, bottom row of Table 5.1. This is reflected in what the Steve and Mary Agnes Hamilton described as the successful *instrumental style* of mentoring.

The instrumental style. Based on the literature from which the TEAM framework is derived, one might expect that the more playing a match did during the first half of their relationship the greater the likelihood of positive outcomes (viz. would correlate negatively with changes in misconduct across four months). This is consistent with the *developmental mentoring* style described by Morrow and Styles (1995). However, two findings counter the expectation that this model would fit best for mentoring youth in the YAP program. First, in Morrow and Styles' study, the mentors and mentees did not come together as a result the youth's misbehavior (their sample was not court-referred youth) or some other agreed upon reason or obvious purpose. Second, their sample was largely children not adolescents.

An alternative candidate for the most effective mentoring style in YAP, and the one on which our second hypothesis was derived, assumes the youth who enter YAP would prefer to engage in the Hamiltons' *instrumental style* of mentoring relationship. In this style, the mentors and mentees move from an initially goal- and skills-development focused approach early in the match toward developing a friendship later in their time together. This style moves the master-and-apprentice ("protégé") relationship toward an instrumental relationship by incorporating a reciprocal friendship once work on an agreed upon task (or problem) has been chose as their focus. Extended to work with youth in the YAP program, we expected that playful interactions later in the relationship would indicate that efforts to foster connection and friendship were underway.

Specifically, play would be more important in the second two-month period because by then the Advocate and youth could really have gotten to know each other, may have done some purposeful problem solving through advocacy efforts, and may now be ready to shift toward friendship development. This pattern of shifting from a focus on the problem or goal, which served as the basis for the formation of their relationship, to a focus on fun activities parallels Hamilton and Hamilton's (2005) *instrumental style* by illustrating a balance over time in focus and purpose. Therefore, using the language of the

TEAM framework, the common adult roles of teacher or coach can expand to a *teacher+* or *coach+* when there is a shift in this balance that introduces a personal relationship and sometimes a friendship that consolidates, forming a mentoring relationship.

This movement from a traditional adult-youth relationship into a coach+, teacher+ or Advocate+ relationship is illustrated in the shift in focus inward to the center column in Table 5.1, which represents an approach that balances the future purpose (left column) and fun purpose (right column). It also tends to happen as a result of an upward shift in focus from a more goal-directed to relational focus, or vice versa, during the relationship.

Another reason to make this our primary hypothesis for this study was that the Hamiltons' research was, like this one, with adolescents (not children), for whom a future focus (requiring overcoming problems with successful engagement in society) is more developmentally salient.

In order to test which of these two styles seemed best in explaining when the YAP mentoring is most effective in lessening misconduct through program participation, we looked at when play occurred. If higher rates of play at the start of the relationship predicted lower misconduct at discharge, the first element of the developmental style would be evidenced. Whereas, if play later in the match was a better predictor, then support for the instrumental style was evidenced.

Both conversations about family, school, friends, and problem-focused discussions were considered to both reflect a more serious, conventional approach to mentoring. Were higher rates of either of these to occur later in the relationship, it could be argued this provided support for the developmental style as most effective. Whereas, if this focus was more predictive of lower misconduct when it happened at the start of the relationship, this would better match the instrumental style of collaborating on a mutually agreed upon, somewhat more future-oriented, problem-resolving, or skill-building focus.

Table 5.2

Data Collection for the TEAM Framework Used Survey Capturing these Activities.

Conventional, (C), Unconventional (U), or Varies (V)	"With your youth/Advocate, how often do you do TALK about each of the following?"	Never	Hardly Ever	Some-times	Often	Very Often
U	Casual Conversation (Discuss sports, what either of you did on the weekend, holiday plans, or other events in town, etc.)	❑	❑	❑	❑	❑
V	Conversation on Social Issues (religion, race, poverty, etc.)	❑	❑	❑	❑	❑
V	Conversation About Relationships: (can check multiple) ❑ Family (C) ❑ Teachers or Employers (C) ❑ Friends, Peers, Peers (U) ❑ Romantic Friend (U)	❑F ❑T ❑Y ❑R	❑F ❑T ❑Y ❑R	❑F ❑T ❑Y ❑R	❑F ❑T ❑Y ❑R	❑F ❑T ❑Y ❑R
V	Listening & Sharing Info about Self (goals, interests, feelings…)	❑	❑	❑	❑	❑
C	Academics (Discuss grades, school, testing…)	❑	❑	❑	❑	❑
C	Behavior (Discuss youth's misbehavior related to problems with peers, teachers, adults, or the courts)	❑	❑	❑	❑	❑
C	Attendance (Discuss importance of showing up school/work)	❑	❑	❑	❑	❑
C	Future Talk (Discuss college, jobs, goals, dreams, etc.; Use computer to research opportunities that would benefit youth.)	❑	❑	❑	❑	❑
U	Play Sports, Athletic Activity, Outdoor Game	❑	❑	❑	❑	❑
U	Creative Activities (do art, read for fun, write a story or song)	❑	❑	❑	❑	❑
U	Played cards, board games, computer games	❑	❑	❑	❑	❑
U	Go to a park, museum, movie, community/cultural event, or college.	❑	❑	❑	❑	❑
C	Get taken somewhere by my Advocate to address a court mandate (e.g., meet with probation officer, meet some curfew)	❑	❑	❑	❑	❑
C	Deal with a crisis or emergency	❑	❑	❑	❑	❑
C	Work on a mandated community service or restitution activity	❑	❑	❑	❑	❑
C	Work on a volunteer community service or "give-back" project	❑	❑	❑	❑	❑

Method of Analyses of Change Over Time

Sample. For the analyses in Study 2, data was available for a total of 161 the matches included in Study 1. The youth were 71% black, 19% Latino, and 12% white, and 68% male. Mentor/Advocate's were 68% black, same sex as their mentees, with education levels ranging from having completed some high school to having earned a college degree.

Data Overview. We used the data we collected a baseline (pre-survey) to get a picture of youths' and Advocates' individual characteristics and starting levels on misconduct at discharge. Then at four months we re-assessed the outcome of misconduct (for youth who had not left YAP by this time point, of course).

At two and four months into the match, which are effectively the middle and end-points for most matches, we asked youth and Advocates to report the frequency at which each of the four kinds of activities (play, casual conversations, problem-related discussions, and "doing" [outside/advocacy efforts]) had happened up to that point in time. Bivariate correlations were examined to estimate the strength of the zero-order relationships between individual characteristics, activity types, and outcomes, specifically misconduct. These provided the foundation for the series of analyses reported below.

Preliminary Tests for Presence of Evidence of Changes in Misconduct over Time

The first step in our series of analyses explaining program-related changes in misconduct was to ensure that pre-post change in misconduct, observed in the t-tests and ANOVAs presented in the prior section, was observed across the three points in time. We wanted to be sure there was linear change or growth over time. Recall that Study 1 reported reductions in the rates of misconduct from pre (entry) to post (discharge) for youth who completed the YAP program but did not assess changes over time within the mentoring relationship. These preliminary analyses were to estimate whether or not there was variability in between these three points in time.

As revealed in 5.1 below, results from a structural model revealed sizeable and statistically significant changes in misconduct between entry and two months as well as between two months and discharge. This change means that the difference (change/slope) and average score (intercept) on misconduct at time four was not explained by change that had already manifest by two months and remained static through discharge. The model reveals there were differences in rates of misconduct reported between entry and two months, as well as between two and four months.

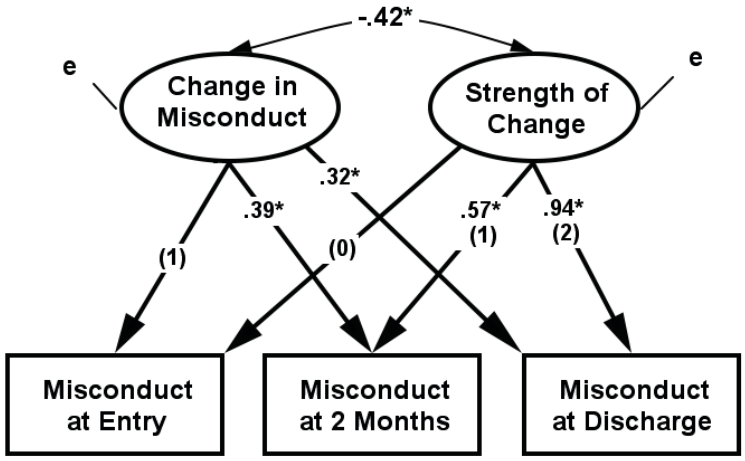


Figure 5.1. Change in Self-Reported Misconduct Across Two Time Intervals

Youth and Advocate Demographic Characteristics

In Table 5.3 below that present correlations between youth and Advocate characteristics, we see that Advocate background experiences, and initial rates of misconduct, were associated with youth characteristics. Starting with levels of initial misconduct, we see that youth sex is the only statistically significant demographic characteristic (of youth or Advocates) that is associated with initial rates of misconduct. It suggests that girls (coded 1) had higher rates of misconduct than boys (coded 0). This variable, therefore was included in all but the first of the models. That way sex-related variability in misconduct was accounted for. The uncommonly higher rates of misconduct among girls than boys is due largely to the fact that the majority of girls came from Vegas, many of whom had been involved in sex trafficking.

Advocate characteristics. Just as there were only weak associations between initial misconduct and types of activities matches engaged in, there were generally weak associations between youth misconduct at entry and their Advocate's background experiences (having been a teacher and highest educational level attained). Youth reporting more misconduct at entry were *not* more likely to receive an older Advocate. Nor were youth reporting more misconduct more likely to have an Advocate with teaching experience. They were, however, slightly more likely to have a more educated Advocate (statistical trend, $p = .065$).

But sex and age of the youths did predict the type of Advocates with whom they were matched. Older youth and girls tended to have more educated mentors. Boys were more likely to have Advocates who were older and who had prior teaching experience. Older youth also tended to have Advocates with teaching experience (statistical trend, $p = .06$). And Advocates with prior teaching experience tended to be older males.

Table 5.3

Correlations Between Demographics, Background Experiences and Initial Misconduct

		Y Age	Y Sex	A Age	A Sex	AP teacher	AP Highest education	Misc Pre
Age	r	1	.095	.161*	.077	.152	.172*	.100
	p		.229	.043	.329	.061	.033	.211
	N	161	161	159	161	154	154	157
Sex	r	.095	1	-.313**	.896**	-.164*	.238**	.212**
	p	.229		.000	.000	.042	.003	.008
	N	161	161	159	161	154	154	157
A2 Age	r	.161*	-.313**	1	-.311**	.208**	-.091	-.106
	p	.043	.000		.000	.010	.260	.188
	N	159	159	159	159	154	154	155
AP teacher	r	.152	-.164*	.208**	-.175*	1	-.056	-.123
	p	.061	.042	.010	.030		.489	.134
	N	154	154	154	154	154	154	150
AP Highest education	r	.172*	.238**	-.091	.221**	-.056	1	.151
	p	.033	.003	.260	.006	.489		.065
	N	154	154	154	154	154	154	150
Misconduct at Entry	r	.100	.212**	-.106	.241**	-.123	.151	1
	p	.211	.008	.188	.002	.134	.065	
	N	157	157	155	157	150	150	157

*= Correlation is significant at the 0.05 level (2-tailed); **= is significant at the 0.01 level.

Sex differences. There were sex differences found in most analyses, starting with the two or three to one ratio of boys to girls. In addition, rates of activities differed at both time points on most types of activities. Table 5.4 below reveals that girls reported a greater frequency of doing things together, engaging in casual talk, and engaging in problem-focused talk than boys in both the first and second half of their relationships. Only on rates of engaging in play did boys and girls not differ.

Table 5.4.
Observed score differences in rates of program activities between boys and girls

		N	Mean	SD	Std. Error	ANOVA	
						F	*p*
Doing	Male	92	3.15	.92558	.09650		
at 2 mos.	Female	35	3.56	.74982	.12674		
	Total	127	3.26	.89742	.07963	5.667	.019
Talk	Male	92	3.55	.80412	.08384		
at 2 mos.	Female	35	3.91	.67412	.11395		
	Total	127	3.65	.78454	.06962	5.487	.021
Play	Male	92	3.35	.87453	.09118		
at 2 mos.	Female	35	3.06	.98769	.16695		
	Total	127	3.27	.91190	.08092	2.448	.120
Problem	Male	92	3.97	.85903	.08956		
Talk	Female	35	4.34	.66682	.11271		
at 2 mos.	Total	127	4.07	.82476	.07319	5.226	.024
Doing	Male	71	3.3056	.99687	.11831		
at 4 mos.	Female	32	3.7500	.99461	.17582		
	Total	103	3.4437	1.01259	.09977	4.389	.039
Talk	Male	71	3.70	.82279	.09765		
at 4 mos.	Female	32	4.17	.55948	.09890		
	Total	103	3.85	.77958	.07681	8.665	.004
Play	Male	71	3.59	.78386	.09303		
at 4 mos.	Female	32	3.53	.96250	.17015		
	Total	103	3.57	.83904	.08267	.109	.742
Problem	Male	71	3.98	.81649	.09690		
Talk	Female	32	4.48	.74151	.13108		
at 4 mos.	Total	103	4.14	.82306	.08110	8.539	.004

Study 2—Analyses Part 1: Observed Activity Scores Predicting Misconduct

The first series of models used observed scores for activities to explain rates of misconduct at discharge from the activity frequency reports, holding constant the role of gender and starting levels of misconduct. As expected based on the sex differences presented in Table 5.4, sex of the youth was a significant predictor of what happened at the start of the match as well as starting levels of misconduct. Girls had higher reported misconduct upon entry than boys, which was somewhat surprising, but reflects the disproportionate number of girls in the Las Vegas YAP, many of whom had been involved in sex crimes. Table 5.5 and Figure 5.2 also suggest the Boys were more likely to begin the matches engaging in playful, recreational activities, while girls were more likely engage first in casual conversations.

There were opposing relationships between the frequency of playing in the first versus second half of the match and reported misconduct at discharge. Although rates of playing were fairly constant across the life of the matches (e.g., $\beta = .50$), higher rates of play *early in the match* predicted higher levels of misconduct at discharge, even after controlling for the contributions of sex and starting levels of misconduct (Table 5.5). Conversely, higher rates of playing *later in the match* predicted lower rates of misconduct at discharge.

That youth in matches that were more playful later in the match tended to report less misconduct after program completion is consistent with the instrumental approach described by the Hamiltons. This assumes, however, that the first half of their relationships were focused on some shared goal-focused activities, like effecting youth-desired changes through advocacy efforts, or spending time learning about each other and identifying shared interests to pursue.

And, indeed in Figure 2 there was a negative association between engaging in the kind of conversations that help each learn about the other (Talk) early in the match and rates of misconduct at discharge. But this path was not statistically significant ($\beta = .14$, $p = .27$). This would have supported the explanation that the more time that was spent getting to know each other at the start of the relationship the lower the rates of misconduct at the end of the relationship. But, again, this association was not statistically different from zero. Perhaps their interaction would have been, but we did not test this.

Therefore, using the casual conversation variable, does not provide additional evidence of an instrumental approach at play, unless the conversation was sufficiently casual to seem to the youth to have no relevance for their future (i.e., conventional, goal-directed focus). Neither do the data support the value of the developmental style, in which play initially serves as the basis for later conversations. Both higher rates of play initially and higher rates of casual conversation (statistical trend at $p = .10$) later were associated with higher rates of misconduct at discharge.

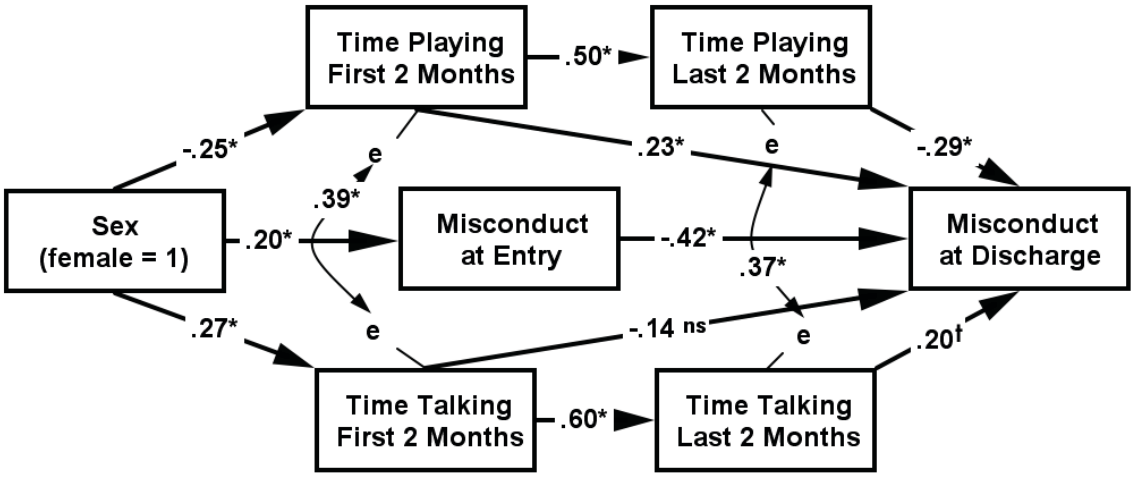


Figure 5.2. Observed score Baseline Model 1: Assessments of talking and playing at two timepoints predicting misconduct at discharge, controlling for youth sex.

Table 5.5

Observed Score Model 1 Fit Indices

```
Number of Free Parameters                       24

Loglikelihood
          H0 Value                       -509.731
          H1 Value                       -502.351

Information Criteria
          Akaike (AIC)                  1067.461
          Bayesian (BIC)                1141.265
          Sample-Size Adjusted BIC      1065.290
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit
          Value                            14.759
          Degrees of Freedom                    9
          P-Value                          0.0978

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                          0.063
          90 Percent C.I.                   0.000  0.119
          Probability RMSEA <= .05          0.308

CFI/TLI   CFI                               0.961
          TLI                               0.910

Chi-Square Test of Model Fit for the Baseline Model
          Value                           169.944
          Degrees of Freedom                   21
          P-Value                          0.0000

SRMR (Standardized Root Mean Square Residual) 0.064

STANDARDIZED MODEL RESULTS
```

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
YPMCS ON				
YPSEX	0.203	0.076	2.675	0.007
Y2PLAY ON				
YPSEX	-0.251	0.082	-3.065	0.002
Y2TALK ON				
YPSEX	0.266	0.081	3.280	0.001
Y4PLAY ON				
Y2PLAY	0.500	0.069	7.217	0.000
Y4TALK ON				
Y2TALK	0.599	0.063	9.486	0.000
Y4MCS ON				
YPMCS	**0.418**	**0.083**	**5.047**	**0.000**
Y2PLAY	**0.232**	**0.106**	**2.185**	**0.029**
Y4PLAY	**-0.289**	**0.108**	**-2.685**	**0.007**
Y2TALK	-0.137	0.121	-1.125	0.261
Y4TALK	**0.202**	**0.122**	**1.656**	**0.098**
Y4PLAY WITH				
Y4TALK	0.372	0.086	4.309	0.000
Y2PLAY WITH				
Y2TALK	0.393	0.075	5.244	0.000

The second model, shown in Figure 5.3 below, incorporates the Advocate characteristics of education level and prior teaching experience into the baseline model presented in Figure 5.3. The first contribution made by Advocate characteristics is through the effect of mentor education on amount of time spent playing later in the relationship (β = -.20). More educated Advocates tended to play less in the second half of the relationship than did less educated Advocates. And because playing later in the relationship predicted lower misconduct at discharge, this relationship reveals a negative indirect effect of Advocates' level of education.

However, there were two positive direct effects of these two mentor characteristics. Mentor education had a direct and positive effect on rates of misconduct at discharge (β = -.22). This model (see 5.6) also revealed a positive effect of having prior teaching experience.). In this way, the youth assigned to the more educated mentors/Advocates and those with more teaching experience seemed to benefit the most from participation in YAP, holding the contribution made by activities constant.

As found in Model 1 without Advocate characteristics, rates of misconduct at discharge vary as a function of when play occurs. In this model, play at both times—early and later in the relationship—make statistically significant contributions to rates of misconduct at discharge. Playing early seems to contribute to more misconduct (β =.26) while playing later contributes to lower rates of misconduct.

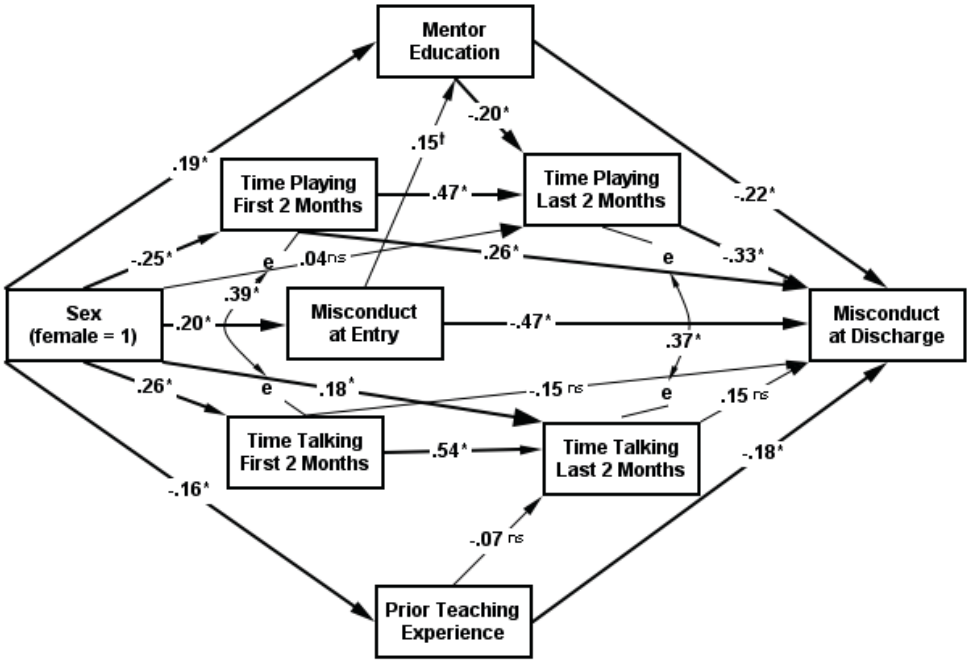


Figure 5.3. Observed Score Model 2: Full Model of Youth, Advocate, and Activity Predictors of Rates of Misconduct at Discharge

Table 5.6
Observed Score Model 2: Fit indices

Number of Free Parameters	37	
Loglikelihood		
H0 Value	-790.889	
H1 Value	-782.711	
Information Criteria		
Akaike (AIC)	1655.778	
Bayesian (BIC)	1769.790	
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	1652.659	
Chi-Square Test of Model Fit		
Value	16.356	
Degrees of Freedom	15	
P-Value	0.3588	
RMSEA (Root Mean Square Error Of Approximation)		
Estimate	0.024	
90 Percent C.I.	0.000	0.080
Probability RMSEA <= .05	0.716	
CFI/TLI		
CFI	0.992	
TLI	0.981	

```
Chi-Square Test of Model Fit for the Baseline Model
          Value                            211.102
          Degrees of Freedom                    36
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual) 0.057

STANDARDIZED MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 YPMCS    ON
    YPSEX              0.201      0.076      2.651      0.008
 MENTORED ON
    YPSEX              0.192      0.078      2.452      0.014
    YPMCS              0.152      0.080      1.900      0.057
 TAUGHT   ON
    YPSEX             -0.160      0.079     -2.023      0.043
 Y2PLAY   ON
    YPSEX             -0.250      0.082     -3.045      0.002
 Y4PLAY   ON
    YPSEX              0.044      0.089      0.496      0.620
    Y2PLAY             0.470      0.076      6.169      0.000
    MENTORED          -0.195      0.084     -2.315      0.021
 Y2TALK   ON
    YPSEX              0.262      0.081      3.219      0.001
Y4TALK   ON
    YPSEX              0.182      0.076      2.382      0.017
    Y2TALK             0.544      0.069      7.917      0.000
    TAUGHT            -0.073      0.067     -1.099      0.272
Y4MCS    ON
    YPMCS              0.467      0.077      6.036      0.000
    MENTORED          -0.224      0.086     -2.624      0.009
    TAUGHT            -0.183      0.076     -2.425      0.015
    Y2PLAY             0.261      0.100      2.610      0.009
    Y4PLAY            -0.334      0.099     -3.360      0.001
    Y2TALK            -0.153      0.114     -1.346      0.178
    Y4TALK             0.154      0.115      1.347      0.178
 Y2TALK   WITH
    Y2PLAY             0.395      0.075      5.292      0.000
 Y4TALK   WITH
    Y4PLAY             0.364      0.088      4.163      0.000
```

While the same contributions made by play in the prior model held, the contributions made by casual conversation were no longer evident. Once the effect of having Advocate with teaching experience was included in the model, the relationships between rates of talking (casual conversations) and misconduct at discharge is no longer statistically significant. It appears having an Advocate with teaching experience may be associated with less talking later in the relationship partly because girls were less likely to have Advocates with teaching experience but engaged in higher rates of casual conversation at both time points. Therefore, the effects of casual conversation may be moderated by sex, and conditional, varying as a function of initial rates of misconduct.

Study 2—Part II: Analyses of Outcomes Related to Factor Score Based Rates of Time Spent Playing and Having Problem-Focused Conversations

This section provides information about how the activities youth and Advocates engage in may contribute to changes in misconduct over the course of treatment using estimates of activities based on factor analyses conducted separately at two and four months. As before, using the TEAM Framework, we attempted a second test of null hypothesis that playful/recreational activities early and problem-focused interactions later, which Morrow and Styles (1995) found worked well early in the match for youth in traditional community based mentoring in the Big Brothers Big Sisters program, also is helpful when working with adjudicated youth. This null was rejected based on the findings about when playful activities were most helpful.

This time, however, we examined specifically whether a match's focus on problem resolution and effecting conventional behavioral changes, which Morrow and Styles found problematic when they persisted across the relationship or when enacted early in the relationship. Both reflect what they labeled the *prescriptive* style. The present analyses examine whether this is similarly problematic when working with youth at much greater risk for criminal activity. Understanding the role of problem-correcting activities in mentoring court-involved youth may help prepared mentors working with other kinds of youth who also are at heightened risk for academic, social, and vocational failure.

It is not uncommon for mentors of high risk youth, like the YAP Advocates, to put a greater emphasis on activities that they believe will serve to eliminate or correct youths' problematic behaviors, attitudes, or knowledge deficiencies; because these problems appear to them, as they do to adults generally, to be what causes these failures.

While that view of how adults help youth is antithetical to the assumptions of many about how mentoring works (Rhodes, 2002; DuBois, Neville, et al., 2002), it is nevertheless one that is pervasive and commonly manifests in adult-youth relationships across the various roles adults play in the lives of youth, from coaches to principals and even parents. Taking an approach based on effecting corrections seems most likely to occur among mentors working with court-involved youth. And, within programs serving court-involved youth, it is possible that the likelihood mentors draw on this prescriptive approach may be even greater for mentees who appear most at risk because of their high rates of misconduct upon entry to the program.

There are two other reasons for the shift to using problem-focused conversations rather than casual conversations in the models using factor scores. First, to confirm that what was happening in the match could be described as more goal directed than friendship-building conversations, we wanted to draw a more clear contrast between talking and problem-solving.

Second, functionally, the CFA model fit indices (see Appendix D) were considerably better for the problem-focused discussions than casual conversations. This

appeared to introduce additional error into estimates, which was what the factor score analyses were intended to correct for (as it is a limitation of observed, mean score variables). Fortunately, running the same model with observed scores for casual conversation and then factor score rates of problem-focused discussion, yielded the same pattern of reveal statistically significant paths.

How perceptions of youth risk may influence the mentors' activity focus. The first hypothesis tested in this section was that youths' starting risk levels, specifically youths' self-reported misconduct at entry into the YAP program, would influence or predict the activities they and their mentors engaged in, especially in early in the mentoring match when mentors may only have this information about youth on which to base their decisions about what their mentees "needs." Given initial risk status was highly related to youth sex, in the tables that follow, factor scores for the four types of activity focuses described in the TEAM framework are presented separately for boys and girls.

The purpose of these tables is to present the general patterns and frequencies of engagement in each activity in the first and second half of the mentoring relationship for each sex with factor scores specifically. For girls, we see that, between the first and second half of the relationship, there is a very small increase in play and similarly a small decrease in problem focused interactions and engagement in the community (Doing). These differences, based on effect sizes, indicate only negligible differences (less than *d* = .05). For boys, only the frequency of problem-focused conversations (*d* = .03) differed over time. But this increase in problem-focused conversations over time did not even approach what is considered a "small" effect (*d* = .20) or difference.

Table 5.7
Activity Frequency Factor Score Descriptive Statistics for Boys

	Minimum	Maximum	Mean	Std. Dev.
Playing & Recreation at 2 Months ("Play2")	-2.13	1.49	.101	.794
Playing & Recreation from 2 to 4 Months ("Play4")	-2.40	1.20	.127	.832
Problem-focus Talk at 2 Months ("Probs2")	-2.87	.89	-.088	.820
Problem-focus Talk from 2 to 4 Months("Probs4")	-2.90	.77	-.133	.730
Casual Conversation at 2 Months ("Talk2")	-1.44	.72	-.067	.455
Casual Conversation from 2 to 4 Months ("Talk4")	-1.31	.49	-.076	.388
Activities in Community at 2 Months ("Do2")	-2.12	1.67	-.115	.888

Activities in Community from 2 to 4 Months (“Do4”)	-2.07	1.43	-.137	.895

Table 5.8
Activity Frequency Factor Score Descriptive Statistics for Girls

	Minimum	Maximum	Mean	Std. Dev
Playing & Recreation at 2 Months (“Play2”)	-2.13	1.49	-.266	.903
Playing & Recreation from 2 to 4 Months (“Play4”)	-2.22	1.20	-.281	1.036
Problem-focus talk at 2 Months (“Probs2”)	-1.52	.89	.232	.646
Problem-focus talk from 2 to 4 Months (“Probs4”)	-2.35	.77	.296	.695
Casual Conversation at 2 Months (“Talk2”)	-.58	.72	.178	.374
Casual Conversation from 2 to 4 Months (“Talk4”)	-.45	.47	.168	.252
Active in Community at 2 Months (“Do2”)	-1.55	1.39	.304	.718
Active in Community from 2 to 4 Months (“Do4”)	-1.52	1.43	.305	.872

Thus, as was the case for the observed scores, the mean levels on factor scores measuring engagement in different kinds of activities did differ by sex, but for youth of both sexes, the average difference between time spent in each type of activity over time was negligible. This continuity in the frequency of each type of activity support the likelihood that differential associations between rates of activities and self-reported misconduct at discharge were more likely due to Advocate characteristics and the nature of the relationship that develops early in the match than just youth characteristic, both age and sex, as well as starting levels of misconduct.

Factor score path models. The factor score path model in Figure 5.4 is the same model tested first using observed scores (Figure 5.2). The fit indices (Table 5.9) suggest this model fit the data well. It reveals that, for the youth as a whole, the level of youth-reported misconduct at entry was not strongly associated with whether the match started with more or less recreation/play ($\beta = .05$) or problem-focused activities ($\beta = .01$). Therefore, what was happening at two months in the match was *not* a function of whether the youth was more or less seen as a (self-reported) trouble maker (i.e., high in misconduct at entry).

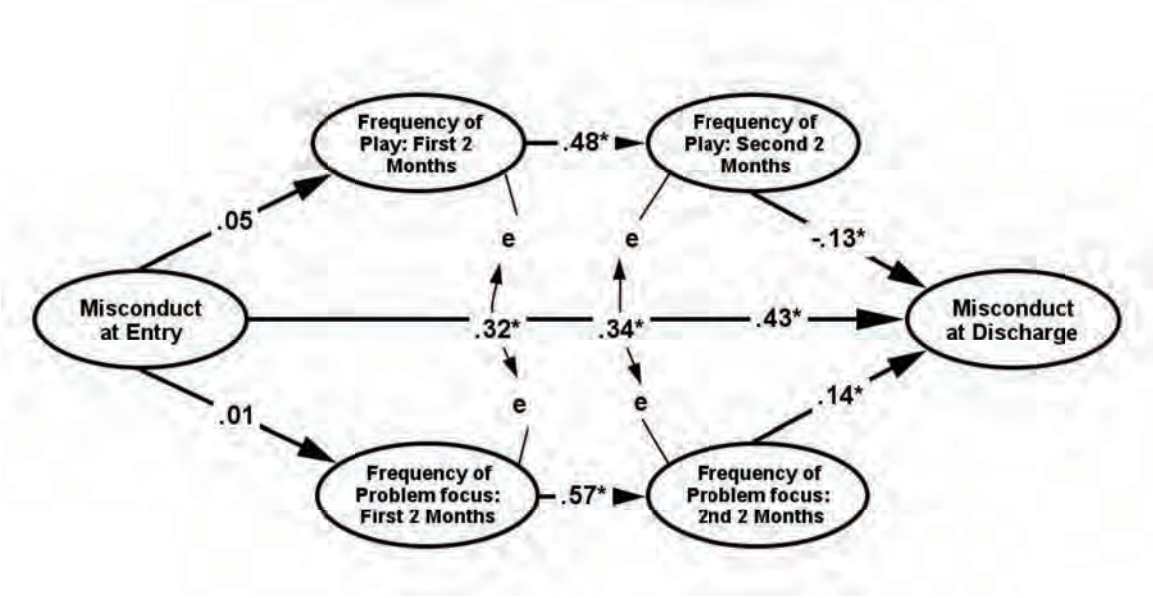


Figure 5.4. Factor Score Model 1: Predicting Misconduct at Discharge from Activities at Two Time Points (Standardized Path Coefficients).

Table 5.9
Fit Indices for Factor Score Model 1: Effects of Activities on Outcomes

```
Number of Free Parameters                       19

Information Criteria
          Akaike (AIC)                  1923.776
          Bayesian (BIC)                1991.502
          Sample-Size Adjusted BIC      1931.264
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit
          Value                            8.074
          Degrees of Freedom                   6
          P-Value                         0.2327

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                         0.036
          90 Percent C.I.                  0.000  0.094
          Probability RMSEA <= .05         0.580

CFI/TLI
          CFI                              0.990
          TLI                              0.975

Chi-Square Test of Model Fit for the Baseline Model
          Value                          222.257
          Degrees of Freedom                  15
          P-Value                         0.0000

SRMR (Standardized Root Mean Square Residual)= 0.034
```

The frequency of play and of problem-focused conversations in the second half of the match were, as with the observed scores, the best predictors of misconduct at discharge. Activity frequencies at two months explained little variability in rates of misconduct at discharge. Rates of misconduct at discharge were lower when more time was spent playing later in the relationship. Here, it should be noted, "playing" still refers to the four types of recreational activities from the activity log in Table 5.2. But the indicator (of the four specific types of activities tallied in the Activity log) that made the greatest contribution to the formation of that factor score later in the match, as can be seen in the factor models presented in Appendix D, was playing sports.

The loadings on the play score varied across assessment periods. For play, the factor for two-month reports of play reflected balanced contributions (factor loadings) of the four play activities although the strongest contributor was sports. At the four-month assessment, however, the factor loading of the sports activity was much larger than the other types of play. The factor loading of play had increased while the other indicators of play decreased. The factor loading for sports was .86 while the factor loadings for the other three types of play ranged from .43 to .51. This suggests that factor scores for rates of play at four months most reflected time spent playing sports.

The contributions to rates of misconduct at discharge made by play and problem-focused conversations were opposite. The findings presented in Figure 5.4 suggest that a higher frequency of play in the second half of the match predicted lower levels of misconduct at discharge (β =-.13, shown as a negative relationship). But, the frequency of time spent focusing on problems in the second half of the relationship (reported at 4 months) predicted higher rates of misconduct at discharge (β =.14, seen as a positive relationship). The effects (or prediction coefficients) for these two activity factors were roughly similar in size and both relatively small. Activity choices early in the relationship were unrelated to outcomes.

The main interpretation of this is that time spent focusing on problems later in the relationship *may* have resulted in increases in self-reported misconduct. These findings are consistent with the instrumental style (described by the Hamiltons) and support hypothesis number two that once a relationship is formed around shared activities, friendship-enhancing activities become more important in influencing outcomes because they consolidate the relationship, shifting it from a coach or instructor adult-youth relationship to a true mentoring relationship that is more comprehensive and balanced in its functions for the youth.

Summary. The types of activities engaged in at two months did not strongly predict the rates of misconduct at discharge, nor were activities at two months related to initial rates of misconduct. This does not support the hypothesis that what happens early in the match is a function of the youths' initial risk level. Nor did initial reports of misconduct at entry predict what occurred later in the second half of the relationship. Higher risk youth did not appear to have relationships that were more problem-focused at

any point in the match.

The second hypothesis, regarding how program outcomes were related to types of activities, also was largely supported by these results. The findings suggest a main, unmediated effect of playful, recreational activities on mentoring outcomes, and negative consequences of focusing on problems rates of misconduct at discharge. This is consistent with the instrumental approach described by the Hamiltons, which is that for teens engaging in activities that offer more shared positive experiences later in the relationship makes a positive contribution to reductions in misconduct, even though rates of play early in the match did not predict rates of misconduct at discharge.

The Role of Advocate Characteristics

In this section we add to the path model above (Figure 5.4) two Advocate characteristics as predictors of activities and of changes in activity frequency over time. Considering the role of Advocate characteristics is useful, in part, because the path model in Figures 5.1-5.4 reveal that once a given type of activity is started, it tends to continue. Beta coefficients in Model 4 predicting the same types of activities happening between two and four months were large for playing ($\beta = .48$) and for problem-focused discussions ($\beta = .58$). This suggests that at least a fourth of the variability in what they were doing later (at four months) is explained by (or could be predicted from) what they were doing earlier (at two months). Advocates education level also was clearly related to activity levels, specifically play. The number of Advocates in each category of educational attainment are listed in Table 5.11 and the average rates of playful activities in the first and second half of the mentoring relationship across these education levels are presented in Figure 5.5, in which each time period is reflected by a separate line.

Table 5.10
Number of Advocates' in Each Category at Each Level of Education Completed

	Level of Education Completed	N
Number of Advocates' at Each Level of Education Completed	1. GED	3
	2. HS graduate	9
	3. Tech school/ apprenticeship	4
	4. Some college	33
	5. Associates degree	13
	6. College graduate	40
	7. Some grad school	7
	8. Grad degree	3

At both assessment points, Figure 5.5 below reveals that youth's reports of time spent playing was higher in matches with Advocates having less education. There is a clear trend toward less play at higher levels of Advocate educational attainment at two ($r = -.07$, $p = .059$) and at four months ($r = -.27$, $p < .00$).

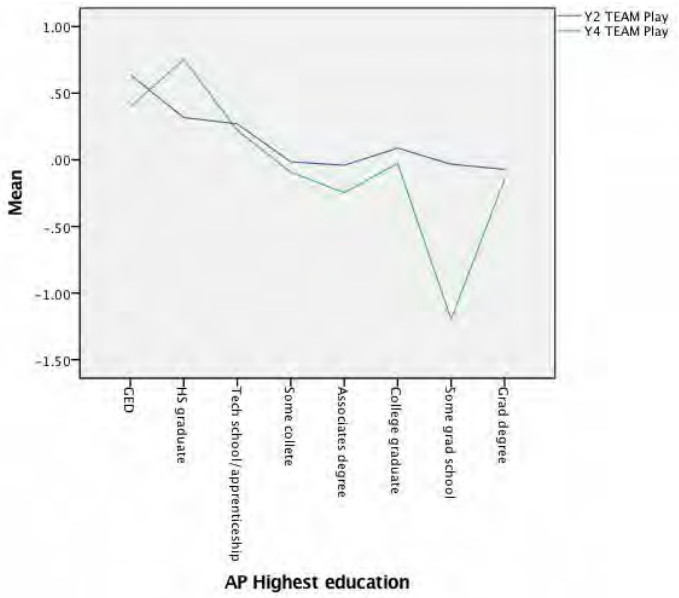


Figure 5.5. Play Frequency at Two and Four Months in Matches Across Advocate's Levels of Education Completed.

The Advocate's level of educational attainment and prior experience as a teacher were included Model 2. Consistent with prior analyses, figure 5.6 illustrates that Advocates with more education tended to report less play in the second half of the relationship. Those Advocates with teaching experience tended to focus less on problem-related conversations later in the relationship.

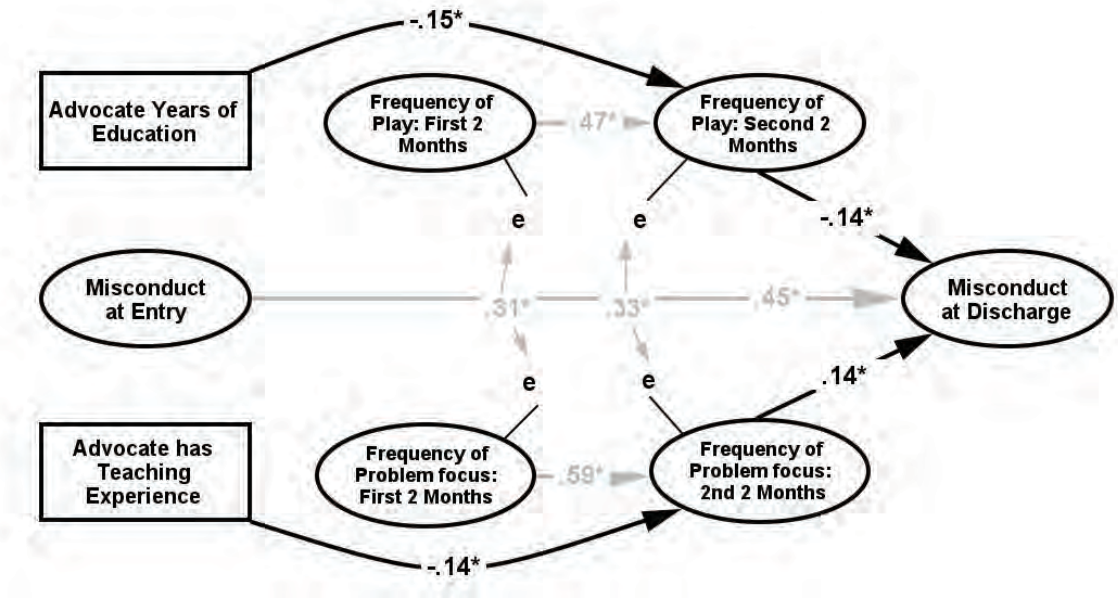


Figure 5.6. Factor Score Model 2: Misconduct at Discharge Predicted From Activities with Indirect Effects of Advocate Backgrounds (with Youth Sex and Age)

What can be observed in Figure 5.6 above is that Advocates with more years of education were less likely to play in the second half of the relationship. Advocates with more teaching experience were *less* likely to engage in problem-focused activities during the second half of the match. Both of these findings are consistent with the instrumental style of mentoring in which a relationship starts off more goal, task, or problem focused but shifts toward the development of a friendship in the later period of the relationship. However, these indirect effects do not include the main or direct effects of these Advocate characteristics observed in the earlier analyses using observed (not factor) scores.

Table 5.11

Fit Indices for Factor Score Model 2: Indirect Effects of Advocate Characteristics

```
Information Criteria
          Akaike (AIC)                    1846.488
          Bayesian (BIC)                  1913.471
          Sample-Size Adjusted BIC        1853.239
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit
          Value                             28.289
          Degrees of Freedom                    16
          P-Value                           0.0292

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.055
          90 Percent C.I.                    0.018  0.088
          Probability RMSEA <= .05           0.360

CFI/TLI   CFI                                0.939
          TLI                                0.932

Chi-Square Test of Model Fit for the Baseline Model
          Value                            220.797
          Degrees of Freedom                    18
          P-Value                          0.0000

SRMR (Standardized Root Mean Square Residual)=0.053

R-SQUARE
    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value
    Y4MCS              0.228      0.052      4.360      0.000
    Y4PLAY             0.239      0.050      4.815      0.000
    Y4PROBLEMS         0.364      0.056      6.517      0.000
```

However, the fit indices for factor score Model 2 in Table 5.11, specifically the CFI/TLI and Chi-Square test are not as good as those in factor score Model 1 without Advocate characteristics (Table 5.9). A reason for this may be that while this model provides additional information about how often different types of Advocates engage in

specific types of activities, there may be other variables not accounted for in this model or main effects of Advocate characteristics not present in the model but are in the data.

Therefore, so far we have examined how Advocate characteristics, specifically their background experiences as teachers and levels of education influence the types of activities that youth engage in with their Advocates. But these indirect tests do not account for whether youth characteristics are systematically associated with the types of Advocates they are assigned or if they make direct or indirect contributions to outcomes.

Two models are presented, discussed, and briefly contrasted below. Each includes the same variables representing these same two Advocate characteristics, but also include the youth demographics of age and sex earlier to be associated with the Advocates' background experiences.

These two models differ in one way. Both are used to explain changes in misconduct as a function of Advocate background characteristics, but the first model considers only the indirect paths (shown in Factor Score Model 3, Figure 5.7, Table 5.12), wherein Advocate background characteristics effect change through their influence on mentoring activities. The second and final model presented below, factor score Model 4 (Figure 5.8, estimates the influence of Advocates' background characteristics both directly and indirectly. The question is, which model fits the data best.

The first model below is consistent with Factor Score Model 2, suggesting that Advocate characteristics, namely prior teaching experience and educational attainment, play a role in the shifts that occur in the types of activities youth and Advocates engage in between earlier and later in their relationship. However, it includes the contributions made by youth age and sex, both of which are associated with Advocate's background characteristics.

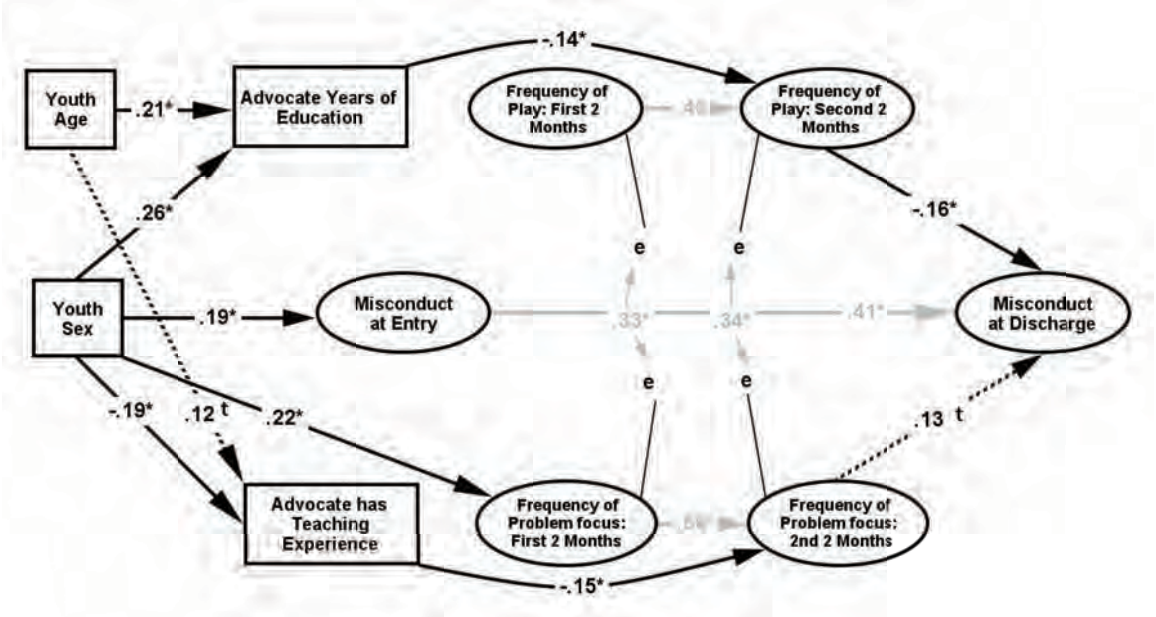


Figure 5.7. Factor Score Model 3: Changes in Misconduct From Activities with Indirect Effects of Advocate Backgrounds and Youth Characteristics

Table 5.12
Fit Indices for Factor Score Model 3: Indirect Effects of Advocate and Youth Characteristics

```
Information Criteria
          Akaike (AIC)                    2956.452
          Bayesian (BIC)                  3066.833
          Sample-Size Adjusted BIC        2968.551
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit
          Value                             44.043
          Degrees of Freedom                    29
          P-Value                           0.0364

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.045
          90 Percent C.I.                    0.012  0.070
          Probability RMSEA <= .05           0.604

CFI/TLI    CFI                                0.944
          TLI                                0.919

Chi-Square Test of Model Fit for the Baseline Model
          Value                            309.386
          Degrees of Freedom                    42
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)=0.057
```

Factor score Model 3 (Figure 5.7) illustrates that the youth's sex matters considerably. There are indirect contributions made by youth age and sex on changes in misconduct as a result of how youth sex and age influence the assignment of Advocates. Yet there are no direct associations between youth characteristics (age and sex) and changes in misconduct or interaction frequency later in the relationship. Therefore, the influence of youth age and sex is indirect, influencing how they get matched with specific types of Advocates, whose background influences the selection of a mentoring approach that is more or less playful or problem-oriented, and consequentially more or less effective, is appears.

Factor score Model 4 (Figure 5.8) tested both direct and indirect effects of Advocate background characteristics on rates of misconduct. In this final Factor Model 4, we see Advocate educational background and prior teaching experience made statistically significant indirect as well as direct contributions to misconduct rates at discharge, when holding constant the other variables in the model (including youth age and sex, starting levels of misconduct, and the frequency of activities early in the match).

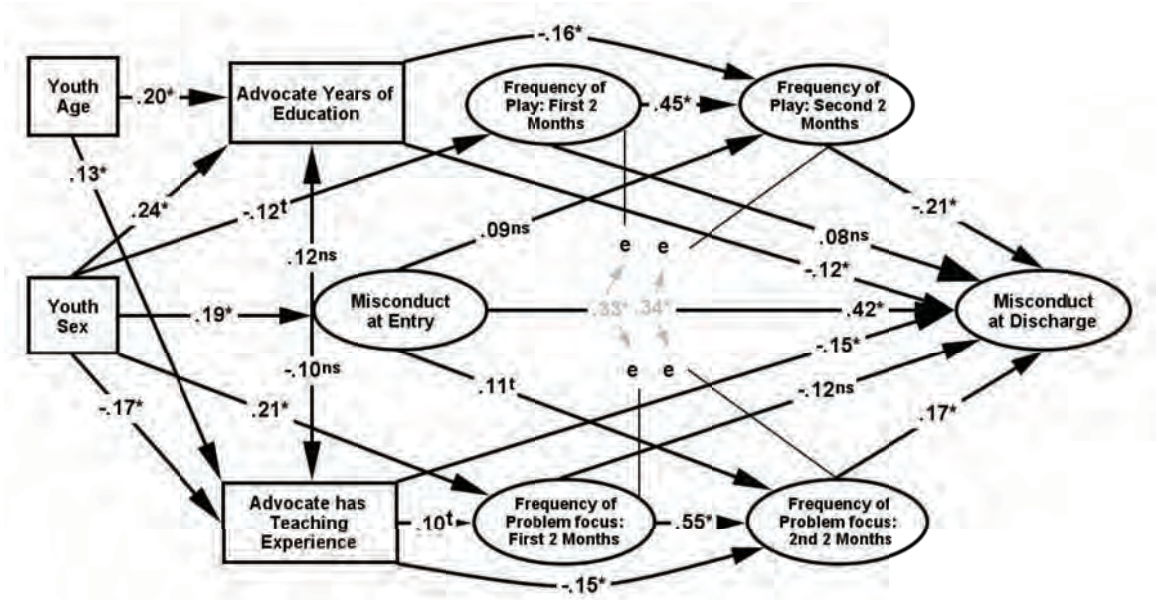


Figure 5.8. Factor Score Model 4: Changes in Misconduct From Activities Including Indirect and Direct Effects of Advocate Background and Youth Characteristics

Advocate education level had a statistically significant positive direct effect on rates of misconduct but a negative indirect effect on the level of misconduct at discharge by contributing to lower levels of play in the second half of the relationship. The direct effect of Advocate education on misconduct is that, after controlling for the influence of several different activities and participant characteristics, having a more educated Advocate predicted lower rates of misconduct, for reasons that are not clear. Yet because more educated mentors played less later in the match, it may be that training them about the benefits of play later in the match may further increase the benefits of having a more educated Advocates serving youth.

Table 5.13
Fit Indices for Factor Score Model 4: Direct and Indirect Effects of Advocate Characteristics

```
MODEL FIT INFORMATION

Number of Free Parameters                       41

Loglikelihood
          H0 Value                       -1433.584
          H1 Value                       -1425.204

Information Criteria
          Akaike (AIC)                    2949.168
          Bayesian (BIC)                  3095.156
          Sample-Size Adjusted BIC        2965.170
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit
          Value                             16.759
          Degrees of Freedom                    19
          P-Value                           0.6062

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.000
          90 Percent C.I.                    0.000  0.047
          Probability RMSEA <= .05           0.962

CFI/TLI   CFI                                1.000
         TLI                                 1.019

Chi-Square Test of Model Fit for the Baseline Model
          Value                            312.718
          Degrees of Freedom                    44
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)  0.034

                                                        Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 YPMCS    ON
    YPSEX              0.188      0.060      3.133      0.002

 APHIEDU  ON
    Y2AGE              0.195      0.058      3.341      0.001
    YPSEX              0.235      0.059      3.994      0.000
    YPMCS              0.118      0.062      1.903      0.057

 APTCHR   ON
    Y2AGE              0.125      0.061      2.054      0.040
    YPSEX             -0.174      0.062     -2.811      0.005
    YPMCS             -0.102      0.067     -1.532      0.126

 Y2PROBLEMS  ON
    YPSEX              0.207      0.062      3.345      0.001
    APTCHR             0.102      0.061      1.670      0.095

 Y2PLAY   ON
    YPSEX             -0.118      0.063     -1.878      0.060

 Y4PLAY   ON
    Y2PLAY             0.453      0.054      8.396      0.000
    APHIEDU           -0.155      0.059     -2.618      0.009
    YPMCS              0.093      0.067      1.383      0.167

 Y4PROBLEMS  ON
    Y2PROBLEMS         0.554      0.049     11.345      0.000
    APTCHR            -0.148      0.052     -2.839      0.005
    YPMCS              0.107      0.061      1.771      0.077
```

Table 5.13 (continued)

Y4MCS ON				
YPMCS	**0.419**	**0.061**	**6.844**	**0.000**
Y4PLAY	**-0.211**	**0.076**	**-2.783**	**0.005**
Y4PROBLEMS	**0.166**	**0.081**	**2.041**	**0.041**
Y2PLAY	0.077	0.075	1.016	0.310
Y2PROBLEMS	-0.118	0.083	-1.419	0.156
APHIEDU	**-0.124**	**0.064**	**-1.940**	**0.052**
APTCHR	**-0.146**	**0.060**	**-2.412**	**0.016**
Y4PROBLEMS WITH				
Y4PLAY	0.336	0.064	5.265	0.000
Y2PROBLEMS WITH				
Y2PLAY	0.317	0.058	5.486	0.000

As in prior models, there was the direct positive effect of having an Advocate with teaching experience. Advocates with teaching experience also had a statistically significant positive indirect effect through their tendency to spent less time having casual conversations later in the match.

Together, these two Advocate background characteristics accounted for three percent of the variance in rates of misconduct reported by youth, holding constant starting levels of misconduct and the other participant background variables in the model. Combined with the frequency of play and casual conversation later in the match, which predicted an additional three percent of variability in changes in misconduct, we see that activities and Advocate characteristics explained over ten percent of the variability in misconduct at discharge.

Conclusions from Study 2

Given the emphasis YAP places on the benefit of similarity between Advocate's and their youth mentees, such as living in the same community or being similar in race or other cultural background characteristics, we examined whether an Advocate's effectiveness varied as a function of his or her level of education or prior teaching experience. These are two cultural background characteristics on which Advocates and their youths' families may differ, and which YAP can control through hiring practices. These variables are also logically related to some of the program related changes on connectedness to teachers, connectedness to school and to educational expectations for success among youth found in Study 1. They also correspond with traditional roles of adults described in the TEAM framework, and having data on the dimensions of a playful focus and mentor-driven goal-directed focus, this study allowed a test of whether these standard roles can be enhanced through the strategic employment of specific activities and specific times in the match.

Advocate education-related characteristics. The findings from this second study do not necessarily comport with the YAP program model assumption that it is those Advocates who are most like their youth—who are most similar to their youth on key background characteristics such as race and familiarity with the youth's community—

who will have the biggest impact. The findings from this study may suggests that there are ways in which differences between youth and Advocate can enhance program effectiveness.

Consistently across models and methods of assessing activity frequency, and most clearly in factor score Model 4, there were direct and indirect benefits of hiring ex-teachers to be Advocates as well as of recruiting more highly educated Advocates. More educated Advocates also may be able to effect more promising outcomes if they pay attention to what they are doing, consider the benefits of incorporating play into their match once it is established (which, again, can be introduced perhaps both through training and supervision), to increase the likelihood that they engage in the kinds of activities that may most benefit the youth.

Activity focus and timing. These analyses suggest that there may be better and worse types of interactions, and that the better mentoring style for YAP participants is an instrumental one. This is partly because the benefits (and iatrogenic effects) of program activities seem dependent on their timing. Playing is something many mentoring programs downplay, and that it seems many adults feel is not a wise use of time with court-involved youth; yet it was a robust albeit modest contributor to program outcomes. Conversely, it does not appear that increasing the frequency at which the match spends time exploring the general aspects of the youths' lives through casual conversation and engaging problem-focused early in the match may maximize the usefulness of this interaction focus.

Two limitations of these analyses and things that are either not known or can't be inferred from these findings should be noted. The first is about causality. These findings, based on correlational not experimental relationships among variables, do not provide definitive evidence that either problem-focused activities or play (independently) *caused* changes in misconduct. What the findings here present are *associations* between the frequency of specific types of activities and subsequent rates of misconduct. Now, the temporal sequence—namely that reports of play frequency are made *before* rates of misconduct are reported—does support the interpretation of this direction of influence. This is because, obviously, rates of misconduct can't cause the frequency of playing in the past. But, reports of misconduct are of how much has occurred in the recent past. So, it is possible that shifts in misconduct happened before the second half of the match and subsequently influenced activities. But that is not the view we are taking of the findings.

Second, although the models suggests that the frequency of activities from mid to end of the relationship are better predictors of rates of misconduct at discharge than the frequency of these activities earlier in the match, this may not be so. It is possible that associations between activities activity frequency reported at four months and changes in misconduct are larger because both were assessed at the same time. Their co-administration may have introduced a method bias that the covariance paths in the models between activities at the different assessment periods did not counter.

Caution also should be used in interpreting these findings, because other important program outcomes, such as 12-month post-discharge educational engagement or most serious disposition, may have quite different associations with program activities and Advocate backgrounds. Similar analyses of the activity-outcome links have not yet been conducted with those long-term outcomes. So we feel it is not advised to move too quickly to effect program policy or hiring changes based on these findings. But this research should prompt those individuals whose approach to mentoring court-referred youth is to employ advocacy as an element of mentoring to consider the value of emphasizing play later in those matches and to try to better understand the reasons for the benefits of play.

A final note about play. What might be reason for the benefits associated with play? On this we can only speculate. A perspective we have given thought to and tried to determine how to test is that play may contribute to declines in misconduct through the way in which fun, playful interactions elicit smiles and positive emotions from mentors. Smiles, which we assume are a logical observable consequence of engaging in play, could convey to the youth a sense of being valued by the mentor. These smiles and positive engagement of the mentors may feel validating, rewarding, and encouraging to the youth; who may subsequently take the relationship more seriously and want to work even harder to prepare for reentry feeling the mentor has become a friend who “has their back.”

APPENDIX A

The creation of narrow age groupings to match treated and untreated youth by age to reduce dependency assumption violations by restricting participants pre-scores (serving as counterfactual untreated comparison group for that age) and treated (post-intervention scores) to only one cohort created the following groupings. This table provides the n for each "group" of treated youth ("how many youth were age ___ after four months of working with an Advocate?") and comparison youth ("how many youth were age ___ before they started working with their Advocate?"). The larger samples ranges in the youngest and oldest ages reflects and effort to address the issue that, by being assessed later, there were more treated responders in the older age cohorts and fewer in the youngest cohorts. So to create somewhat more balanced groups, a wider age range was used. These also restricted participants' responses to only one age cohort, which resulted in the exclusion of some participant data to allow statistically independent comparisons in these groups.

Group	Treated Group	Untreated Group
1 (11-12 years)	5	6
2 (13yr 0mo – 13yr 5mo)	2	5
3 (13yr 6mo – 13yr 11mo)	5	8
4 (14yr 0mo – 2mo)	3	3
5 (14yr 3mo – 5mo)	6	8
6 (14yr 6mo – 8mo)	8	7
7 (14yr 9mo – 11mo)	6	4
8 (15yr 0mo – 2mo)	5	7
9 (15yr 3mo – 5mo)	3	11
10 (15yr 6mo – 8mo)	12	11
11 (15yr 9mo – 11mo)	13	12
12 (16yr 0mo – 2mo)	10	8
13 (16yr 3mo – 5mo)	10	10
14 (16yr 6mo – 8mo)	8	17
15 (16yr 9mo – 11mo)	17	10
16 (17yr 0mo – 2mo)	12	13
17 (17yr 3mo – 5mo)	13	8
18 (17yr 6mo – 8mo)	5	7
19 (17yr 9mo – 11mo)	8	3
20 (18yr 0mo – 5mo)	11	6
21 (18yr 6mo – 11mo)	2	0

These comparisons are of limited use because small sample size of each cohort that undoubtedly generated unstable, inaccurate mean scores with huge variability. Below are select results from the series of t-tests conducted to compare outcome variables across the treatment and untreated comparison groups within each narrow (3 mos.) age cohort grouping category to estimate treatment effects across 4 mos.. Those variables that had significant and near-significant changes are listed below for each outcome variable by age-cohort group. These are purely informational and should not be interpreted as reliable estimates (none meet needed power requirements to conduct these significance tests and the small sample sizes tended to yield non-normal distributions), but purely to convey some information about in what age cohorts changes related to program participation seem more likely and for others.

Running only t-tests (with a p-value of .10 used as evidence of non-negligible differences), the analyses below suggest that future expectations more often differed between untreated and treated youth than did rates of self-reported misconduct and connectedness. But the

patterns of change were equivocal, except for expectations to finish high school and self-reported misconduct which favored the program completers. The group with the higher score is in bold.

Outcome variables on which the treated and untreated differed	n for treated & untreated	Treated mean	Untreated mean	diff	T-test	*p*
Expect to Finish High School						
Group 2 (13 yr 0 - 5mo)	2, 5	**4.67**	3.20	1.47	2.34	.067
Group 8 (15yr 0 – 2mo)	5, 7	**3.97**	3.23	.733	1.89	.064
Group 14 (16yr 6 – 8mo)	8, 17	**3.91**	3.59	.32	1.17	.102
Group 15 (16yr – 11mo)	17, 10	**3.66**	3.00	.66	2.28	.031
Group 16 (17yr 0 – 2mo)	12, 13	3.22	**3.64**	-.421	-1.87	.075
Expect to Start College						
Group 3 (13yr 6 - 11mo)	5, 8	2.90	**3.73**	-.83	-2.68	.022
Group 5 (14yr 3 – 5mo)	6, 8	**3.70**	2.47	1.23	2.98	.012
Group 10 (15yr 6 – 8mo)	12, 11	3.33	**2.55**	.78	1.81	.091
Group 15 (16yr – 11mo)	17, 10	**3.12**	2.00	1.18	3.01	.006
Expect to Finish College						
Group 3 (13yr 6 - 11mo)	5, 8	2.88	**3.61**	-.73	-1.85	.092
Group 5 (14yr 3 – 5mo)	6, 8	**3.99**	2.67	1.33	3.36	.010
Group 6 (14yr 6 – 8mo)	8, 7	3.10	**3.14**	.86	2.12	.079
Group 10 (15yr 6 – 8mo)	12, 11	3.46	**2.55**	.92	2.06	.056
Group 15 (16yr – 11mo)	17, 10	**3.06**	2.10	.955	2.14	.043
Expect to Find a Job						
Group 10 (15yr 6 – 8mo)	12, 11	**3.51**	2.93	.57	1.96	.064
Group 16 (17yr 0 – 2mo)	12, 13	3.16	**3.62**	-.45	-2.09	.048
Group 20 (18yr 0 – 5mo)	11, 6	**3.84**	3.05	.79	3.38	.004
Expect to Succeed at Job						
Group 10 (15yr 6 – 8mo)	12, 11	**3.53**	2.91	.62	2.19	.047
Group 16 (17yr 0 – 2mo)	12, 13	3.28	**3.85**	-.57	-3.30	.003
Group 19 (17yr 9 – 11mo)	8, 3	3.54	**4.00**	-.46	-2.30	.041
Group 20 (18yr 0 – 5mo)	11, 6	**3.76**	3.12	.65	3.10	.007
Expect to stay out of court syst.						
Group 7 (14yr 9 – 11mo)	8, 7	3.02	**4.00**	-9.80	-2.45	.058
Group 9 (15yr 3 – 5mo)	3, 11	**4.00**	3.62	.38	1.87	.091
Group 11 (15yr 9 – 11mo)	13, 12	3.16	**3.75**	-.59	-2.04	.053
Connectedness to Family						
Group 9 (15yr 3 – 5mo)	3, 11	-.42	**.26**	-.68	-1.81	.096
Group 10 (15yr 6 – 8mo)	12, 11	**.44**	-.12	.56	1.87	.076
Connectedness to School						
Group 11 (15yr 9 – 11mo)	13, 12	-.43	**.18**	-.60	-1.90	.072
Group 13 (16yr 3 – 5mo)	10, 10	**.21**	-.32	.52	1.76	.096
Group 15 (16yr – 11mo)	17, 10	**.11**	-.37	.48	2.47	.021
Connectedness to Teacher						
Group 11 (15yr 9 – 11mo)	13, 12	-.39	**.26**	-.65	-1.79	.087
Connectedness to Self-in-Future						
Group 11 (15yr 9 – 11mo)	13, 12	-.35	**.10**	-.45	-1.98	.060
Group 19 (17yr 9 – 11mo)	8, 3	.03	**.44**	-.41	-2.13	.062
Misconduct (lower mean bolded)						
Group 3 (13yr 6 - 11mo)	5, 8	**.12**	-.39	.52	2.56	. 027
Group 11 (15yr 9 – 11mo)	13, 12	**.22**	-.20	.41	2.26	.034

APPENDIX B.
Piloting or Model Fitting Items in Five Advocacy Measures

Five measures of advocacy were created, replicated or tested in this study and exploratory associations with Advocate characteristics are reported in this Appendix. Without sufficient validity evidence—relying only on content and factorial validity evidence provided later in the Appendix—these analyses should be considered exploratory. First, we described the nature and initial development of each of these measures.

DuBois Advocacy, Instruction, and Reflecting (D-AIR) scale

A set of items developed by DuBois to assess advocacy and other mentoring activities were factor analyzed to create one measure of advocacy as a part of an approach to mentoring youth. These items capture youths' perceptions of things the Advocate does for the youth, mostly when the Advocate is with the youth. These activities balance problem solving, role modeling, interpersonal sharing, helping with feelings and trying to understand and respond to the needs of the youth.

In DuBois et al. (2013) meta-analysis, they reported that programs emphasizing advocacy and/or teaching had larger effects that those that did not. This measure, similarly, collapses advocacy and teaching but also includes indicators of role modeling and self-reflection coaching. This view of advocacy is akin to an instrumental mentor whose initial focus is on problem-focused discussions, doing, and mastering to help the youth in becoming something or someone.

The three main activities represented in what we are calling the D-AIR measure appear to be advocacy (items 8, 9 and 10), instruction (items 1, 2, 3, 4), and reflection (items 5, 6, and 7). These items held together as one factor (see Appendix) such that each contributed somewhat similarly to the one overall approach to mentoring. We assigned the acronym D-AIR to reflect the source of the items (DuBois) and the three functions it captures: advocacy (A), instruction (I), and reflection (R).

1. says or does something to help me with my feelings
2. gives me advice or ideas about how to do something
3. shows me how to do something
4. helps me practice something
5. shows or tells me things about his/her own life
6. helps me think about myself or the world in a different way
7. helps me figure out what is really important to me
8. *helps me by talking with other people who are important in my life*
9. *makes sure I have the things I need to be successful*
10. *makes sure I have a chance to participate in activities that are good for me*

The Blue Ribbon Advocacy Scale

The Blue Ribbon Scale was developed based on a definition of advocacy made by Graig Meyers, then Executive Director of the Blue Ribbon Advocacy Program. His definition generated six items. Through a series of exploratory and confirmatory factor analyses, a scale was created and the scale ended up including four of these items. The two that were dropped either double-loaded on two factors two factors. To create distinct measures, items that loaded on the Blue Ribbon factor and the D-AIR factor were eliminated. For this reason, the D-AIR scale also had to drop the last two items (11 and 12) that were redundant or cross-loaded with items in the Blue Ribbon scale.

The four items on this scale are all about trying to effect change in the youth's world and trying to bridge the worlds of work, school, family, and mental health staff. This is advocacy as outreach and fostering collaborations, as compared to D-AIR focus which is more interpersonal, psychosocial, skill-building advocacy assessment.

1. goes to parent-teacher conferences or meetings (originally item #2)
2. talks with my teacher/teacher aide (originally item #3)
3. talks with my therapist/counselor (originally item #4)
4. communicates with people to help youth get a job, get into school, a training program (originally item #5)

YAP Advocacy Scales from the PPV study of YAP lead by Herrera

An earlier study of the YAP program led by Carla Herrera generated a set of items that were direct representations of what Advocates tended to do with youth. We conducted additional analyses and confirmatory factor analyses to estimate several types of Advocacy common in YAP (see Appendix). These are helping youth develop self-management skills, advance vocationally, develop connections in the community, and assist with family issues. We focused on the first three of these in some preliminary analyses.

Individual Self Management Adv. Focus (Relational)(alpha = .74)

1. Worked with individually on life skills such, as healthy decision-making, dealing with stress and avoiding risky behaviors?
2. Worked with individually to help him/her manage anger more appropriately?
3. Worked with individually to help him/her avoid using drugs or alcohol?

Individual Vocational Advocacy Focus (Goal directed))(alpha = .80)

1. Helped with job searches, job applications, resumes or interviews?
2. Helped find a job (either a supported work job or a regular paying job)
3. Talked to about appropriate work place dress and behavior?
4. Worked with individually on independent living skills such as budgeting, time management or using public transportation?

Individual Community Advocacy Focus (Relational)(alpha = .77)

1. Introduced to an adult in the community who could be a mentor, role model or friend to the youth?
2. Introduced to another youth in the community who could be a role model or friend to the youth?
3. Helped enroll the youth in an arts, sports or recreation club, program or center in the community?

Advocate Personality Characteristics and Advocacy Approaches

To better understand what types of Advocates more often engage in interpersonal mentoring or ecologically based advocacy in the community, we examined the associations between three personality characteristics and these advocacy scales. Using Exploratory factor analyses on the Advocate pre-match (before meeting the youth) data, with a rotated component matrix, the factor loadings and reliabilities for the Extraversion, Neuroticism, and Agreeableness personality scales are listed below. Item loadings on other scales are included below (but not on computed subscale scores) to illustrate overlap.

	Factor		
Stem: I see myself as someone who…	1	2	3
Extraversion			
...is talkative.	**.832**		
...is outgoing, sociable.	**.820**		
...tends to be quiet. (R)	**.749**		-.234
...is enthusiastic.	**.718**		.298
...is full of energy.	**.717**		
...has an assertive personality.	**.596**		
...is sometimes shy. (R)	**.531**	-.388	-.265
Neuroticism			
...can be tense.		**.734**	
...gets nervous easily.	-.312	**.704**	
...worries a lot.		**.694**	
…is easily upset.		**.656**	-.224
…can be moody.		**.638**	
...is relaxed, handles stress well. (R)	-.221	**.616**	-.262
...remains calm in tense situations. (R)		**.580**	
Agreeableness			
...is kind and considerate to almost everyone.			**.794**
...is helpful and unselfish with others.			**.738**
...likes to cooperate with others.	.401		**.627**
...has a forgiving nature.			**.606**
...is sometimes rude to others. (R)	-.303	-.300	**.517**
...is generally trusting.	.219		**.508**
...starts quarrels with others. (R)			**.424**
Coefficient a	.85	.80	.74

Items for loadings less than .2 were been omitted. Bolded items within the same column are included in the same scale. Items denoted by (R) were reverse scored. The metric for these scales was from 1 (strongly disagree) to 5 (strongly agree) with 3 (neither agree nor disagree) as the midpoint.

Variances: Factor 1: "Extraversion" = 22.472
Factor 2: "Neuroticism" = 15.264
Factor 3: "Agreeableness" = 13.181

Big 5 personality characteristics. We were also curious about how Advocate personality characteristics might predict the types of activities Advocates engaged in with youth. The three personality characteristics we examined were extraversion, neuroticism, and agreeableness. (The items included in the three personality scales used to generate factor scores are provided in the Appendix.)

Extraversion. Extraverted Advocates described themselves as more talkative, outgoing/sociable, enthusiastic, full of energy, and as having an assertive personality, tending not to be quiet, but only sometimes shy.

Neuroticism. Advocates high on the Neuroticism scale said about themselves that they can be tense, get nervous easily, worry a lot, are easily upset, can be moody and are not relaxed, don't handle stress well, and do not remain calm in tense situations.

Agreeableness. Advocates high on the Agreeableness scale tended to more often report seeing themselves as kind and considerate to almost everyone, helpful and unselfish with others, cooperative with others, forgiving in nature, not often rude to others, generally trusting, and so rarely quarrel with others.

In the correlation table that follows, the top row lists the scales, and the two in grey are the two Advocacy scales developed for use in this study. In addition to the three personality characteristics of extraversion, neuroticism, and agreeableness, the three other advocacy approaches that specifically reflect what Advocates in YAP do (generated in a prior PPV study) also are included (The items and factor analyses conducted to determine which items best reflected each construct are provided in a later Appendix.)

Personality Characteristics as Predictors of Self-Reported Advocacy Activities

	Blue Ribbon (Advocate report at two months, A2) Advocacy	D-AIR Advocacy, Instruction, & Relating Scale (A2)	Advocate Extra-version	Advocate Neurotic-ism	Advocate Agree-ableness
Advocate D-AIR at 2 months	.20*				
AP Extraversion	.105	.23**			
AP Neuroticism	-.103	-.22*	-.50**		
AP Agreeableness	.073	**.048**	.43**	-.45**	
A2 YAP Individual Self management Adv.	.204*	.578**	.113	-.159✚	.113
A2 YAP Individual Vocational Advocacy	.247**	.319**	.140	-.252**	.189*
A2 YAP Individual Community Advocacy	.424**	.359**	.068	-.158	.028

When we consider the DuBois Advocacy scale in contrast to the Blue Ribbon Scale, the correlations suggest those who engaged in higher levels of the advocacy, instruction, and reflection (D-AIR) were often those reporting high levels of extraversion and low levels of neuroticism. This is consistent with the assumption that someone taking an interpersonal, reflective, teaching, personally disclosing approach would be more extraverted and have a thicker skin, particularly with kids who are guarded emotionally. When correlated with the three YAP Advocacy scales, we found Advocates taking the D-AIR, inclusive approach to advocacy-as-mentoring, also tended to focus on self-development advocacy, helping youth make better decisions, address substance use and anger issues, and cope with problems. Slightly less strong were associations between the D-AIR approach and the YAP career and community advocacy scales. But, when comparing the D-AIR and the Blue Ribbon scale, the greatest distinction is the co-occurrence of the D-AIR approach and the YAP advocacy as assisting in self-management approach.

The Blue Ribbon Advocacy Scale correlated only moderately (r = .20) with the D-AIR and was most strongly related to the YAP community advocacy scale, which makes sense given both are about activities done in order to strengthen the ties across the

youths' social ecologies (school, work, clubs). Those who frequently engaged in the Blue Ribbon scale advocacy activities of direct outreach to others to explicitly Advocate for the youth or help navigate the youths' efforts to be successful in school, work, family or social settings also tended to engage in interpersonal discussions of risk-taking (drugs), decision-making, and feelings (self-management), but the association between these was weaker. Those activities require more one-on-one interactions with the youth, which are more similar to the approach assessed by the D-AIR.

Both DuBois' and the Blue Ribbon Advocacy Scale were associated with the YAP Vocational Advocacy scale. DuBois' scale taps into the part of vocational advocacy that is about helping the youth learn about the world, about jobs, about how to prepare to look for a job and helping youth think about what type of work they might like to do. The Blue Ribbon scale (similar to items on the YAP Vocational Advocacy scale) reflects actually interfacing with people in the world who can help the youth secure a job.

APPENDIX C

Factor Models and Fit Indices for the TEAM constructs: Play, Talk, Learn, and Do Item Listing and Associated Confirmatory Factor Analyses

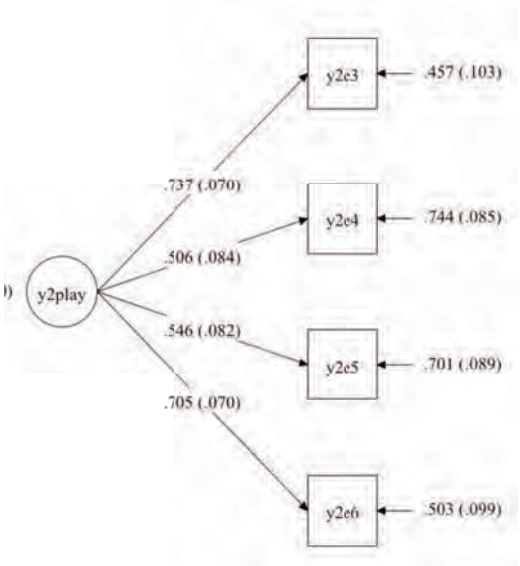
Item Indicator Shown in Factor Models	"With your youth/Advocate, how often do you do TALK about each of the following?"
D1	**Talk:** Casual Conversation (Discuss sports, what either of you did on the weekend, holiday plans, or other events in town, etc.)
D2	**Talk:** Conversation on Social Issues (religion, race, poverty, etc.)
D3	**Talk:** Conversation About Relationships: (can check multiple) ❑ Family (f) ❑ Teachers or Employers (te) ❑ Friends, Peers, Peers (y) ❑ Romantic Friend (r)
D4	**Talk:** Listening & Sharing Info about Self (goals, interests, feelings…)
D5	**Learn:** Academics (Discuss grades, school, testing…)
D6	**Learn:** Behavior (Discuss youth's misbehavior related to problems with peers, teachers, adults, or the courts)
D7	**Learn:** Attendance (Discuss importance of showing up school/work)
D8	**Learn:** Future Talk (Discuss college, Jobs, goals, dreams, etc.; Use computer to research opportunities that would benefit youth.)
C3	**Play:** Play Sports, Athletic Activity, Outdoor Game
C4	**Play:** Creative Activities (do art, read for fun, write a story or song)
C5	**Play:** Played cards, board games, computer games
C6	**Play:** Go to a park, museum, movie, community/cultural event, or college.
E1	**Do:** Participate in a group class (e.g., life skills, anger management)
E2	**Do:** Homework/Tutoring (reading, academics, computer or library work)
E7	**Do:** Get taken somewhere by my Advocate to address a court mandate (e.g., meet with probation officer, meet some curfew)
E8	**Do:** Deal with a crisis or emergency
E9	**Do:** Work on a mandated community service or restitution activity
E10	**Do:** Work on a volunteer community service or "give-back" project

In the factor models that follow, TEAM refers to the four activity construction (Playing, Talking, Learning/Problem Correcting and Doing), Y is for youth report, A is for Advocate report, 2 is for assessment at 2 months, and 4 is for assessment at 4 months. "Mod. indices" refers to modification indices proposed by Mplus (the statistical program) to increase model fit.

APPENDIX D: Confirmatory Factor Models for Activity Frequency Rates

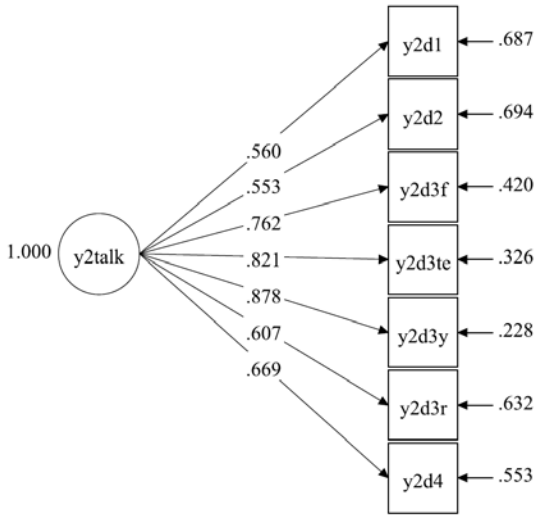
Y2 Play

Chi square 2.722, df = 2, p = .256, chi square/df = 1.361
RMSEA .053, CI .000-.192, p < .05 = .361
CFI .992 SRMR = .027
All fit indices suggest a good model fit. No mod. indices suggested or employed.



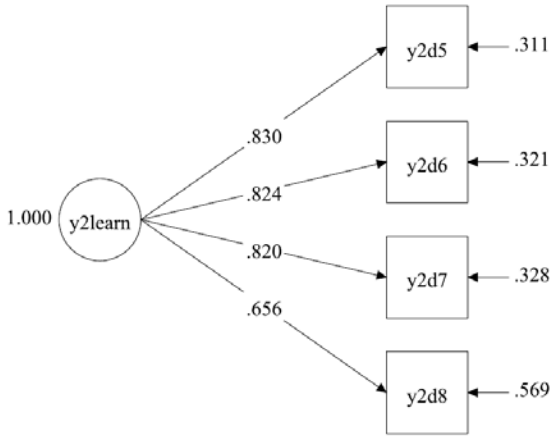
Y2 Talk: Casual Conversations

Chi square = 42.914, df 14, p = .0001, chi square/df = 3.065
RMSEA = .128, CI .085-.172, p < .05 = .002
CFI = .918 SRMR = .056
SRMR is only adequate fit index. No mod. indices suggested or employed.



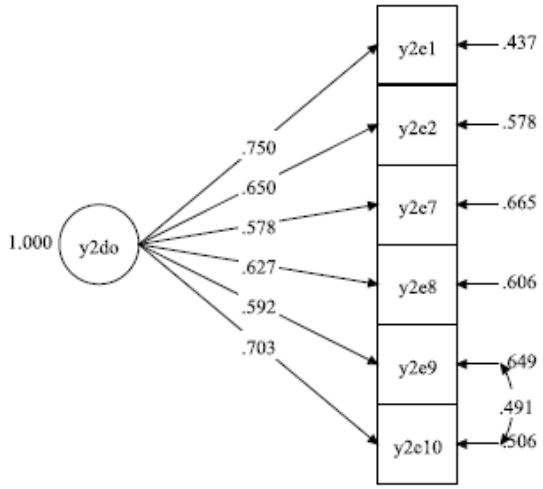
Y2 Problem-focused Discussions

EFA suggested items loaded onto one factor. Alpha was .862 and did not improve by removing items.
Chi square = 10.737, df 2, p = .0047, chi square/df = 5.369
RMSEA = .185, CI .088-.301, p < .05 = .015
CFI = .964, SRMR = .030
CFI and SRMR suggest good model fit. No mod. indices suggested or employed.



Y2 Do
EFA suggested one factor. Observed score alpha was .830 and did not improve by removing any items. The CFA yielded a poor model fit, but mod indices suggested 9 with 10 (items about comm. service), which improved the model.

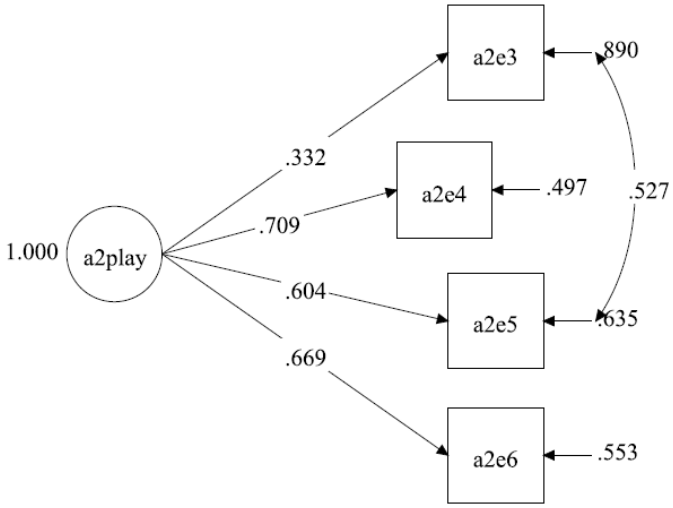
Chi square = 9.854, df = 8, p = .275, chi square/df = 1.232
RMSEA = .043, CI = .000-.118, p < .05 = .491
CFI .993 SRMR = .030
All fit indices suggest a good model fit. No mod. indices suggested or employed.



Advocate reports

A2 Play

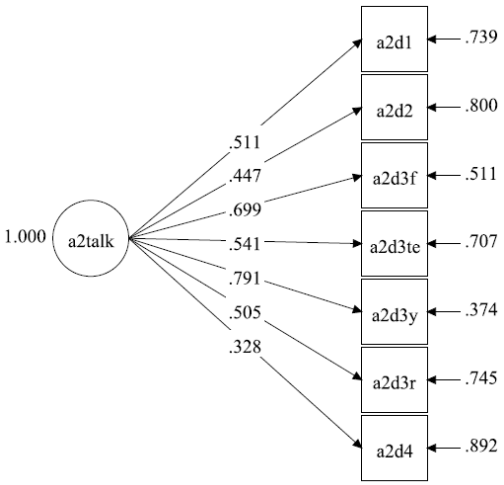
Chi square = 2.667, df 1, p = .1024, chi square/df = 2.667
RMSEA = .111, CI .000-.282, p < .05 = .160
CFI = .987 SRMR = .023
Chi square, CFI and SRMR suggest a good model fit. No mod. indices suggested or employed.



The item E3 (played sports, athletic activities, or outdoor games) does not really contribute to the factor after accounting for the variance it shares with item E5 (Played card, board games, computer games). This pattern is the opposite pattern observed for youth reports, especially at four months.

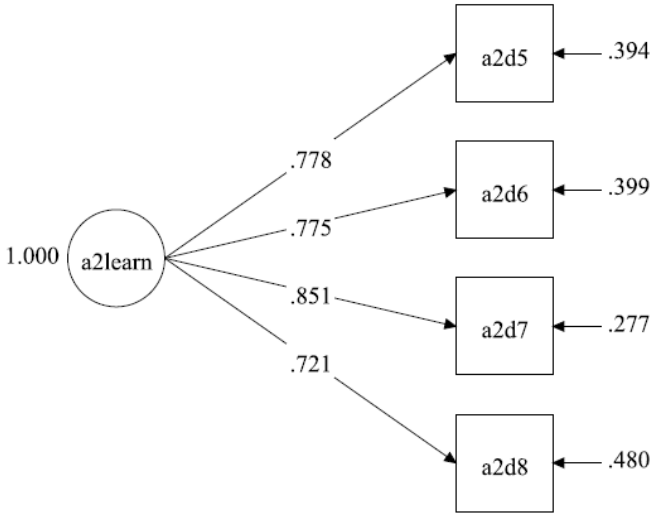
A2 Talk

Chi square 22.233, df 14, p = .0739
RMSEA = .066, CI = .000-.115, p < .05 = .274
CFI = .951 SRMR = .046 No mod. indices suggested or employed.



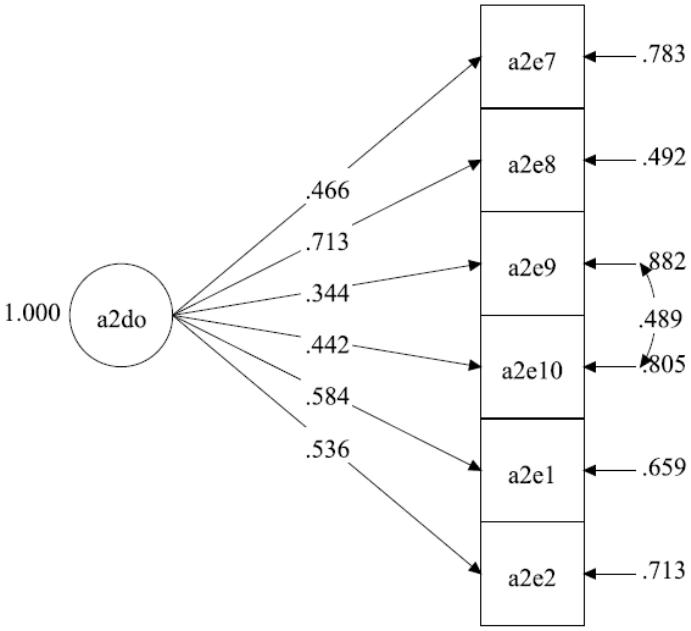
A2 Problem-focused Discussions

Chi square = .972, df 2, p = .615, chi square/df = .486
RMSEA = .000, CI = .00-.138, p < .05 = .704
CFI = 1 SRMR = .010
All fit indices suggest a very good model fit. No mod. indices suggested or employed.



A2 Do

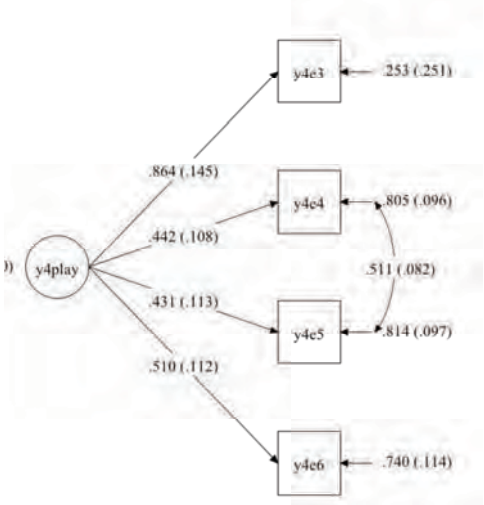
Chi square = 7.888, df 8, p = .4444, chi square/df = .986
RMSEA = .000, CI .000-1.00, p < .050 .668
CFI 1.0 SRMR .034
Note that "dealing with crisis or emergency" (E8) is largest contributor to latent variable.



Youth Report at 4 Months

Y4 Play

Initial fit was poor, but improved with MIs suggested Y4e5 WITH Y4c4.
Chi square = 12.70, df 2, p = .6886, chi square/df .161
RMSEA = .000, CI .000-.193, p < .05 = .724
CFI = 1.0 SRMR .006
Fit indices suggest a good model fit. No further mod. indices suggested or employed.



Y4 Talk: Casual Conversation

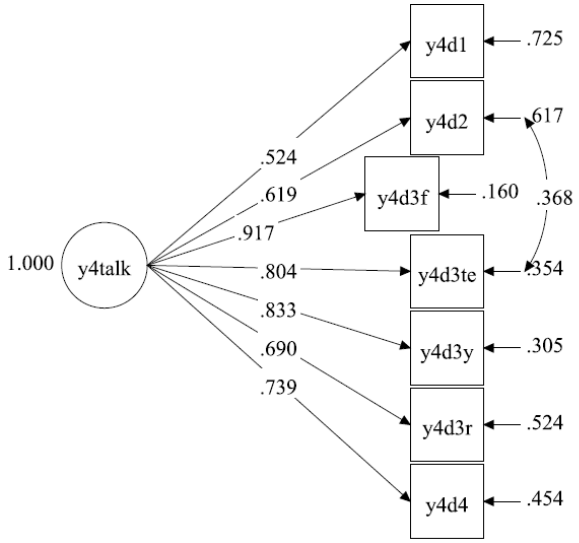
In the EFA, Casual Conversation items loaded onto two factors. Suggested not all topics of conversation occurred or co-occurred. The same issue appeared in the CFA models.

Rotated Component Matrix

	Component	
	1	2
A4 TEAM Talk 1	.182	.790
A4 TEAM Talk 2	-.050	.685
A4 TEAM Talk 3 Family	.537	.657
A4 TEAM Talk 3 Teach/Employer	.826	.327
A4 TEAM Talk 3 Yth	.800	.300
A4 TEAM Talk 3 Romantic	.814	-.073
A4 TEAM Talk 4	.315	.639

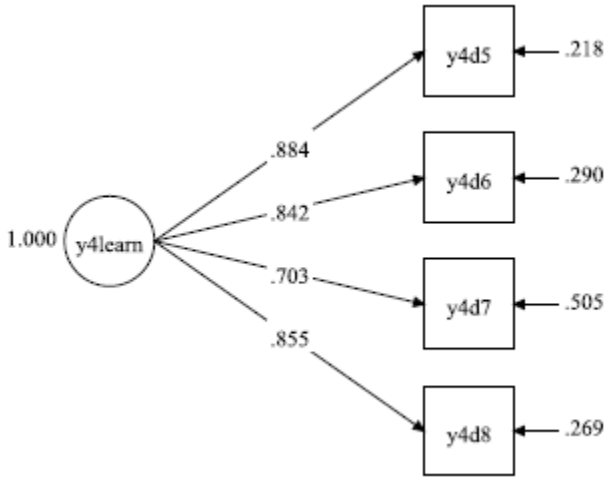
Y4 Talk: Casual Conversation

Chi square = 22.911, df 13, p .0428,
Chi square/df = 1.762
RMSEA .086, CI = .015-.143, $p < .05 = .146$
CFI = .971 SRMR .036



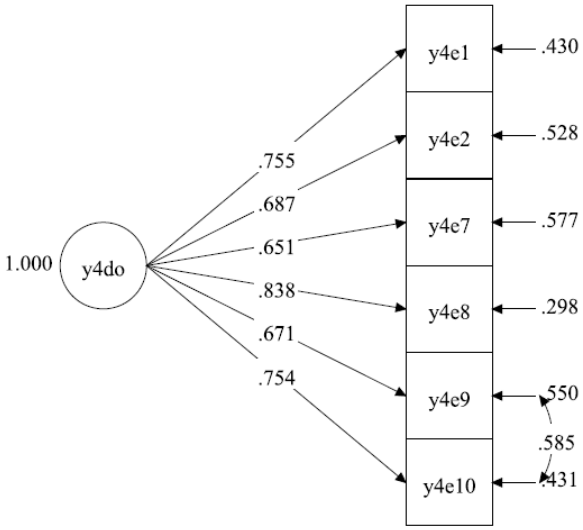
Y4 Problem-focused Discussions

Chi square = 5.195, df 2, p = .0745, chi square/df = 2.5975
RMSEA = .125, CI .000-.261, p <.05 = .126
CFI = .987 SRMR = .019 No mod. indices suggested or employed.



Y4 Doing in the Community

EFA items loaded on factor. Alpha = .886 for observed score assessment.
Fit indices suggested a poor model fit. Mod indices suggested adding covariances between Y4E10 WITH Y4E9



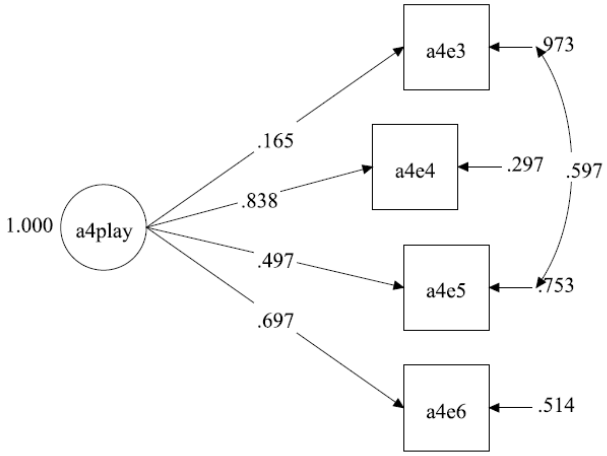
Revised model:
Chi square = 22.251, df 8, p = .0045, Chi square/df = 2.781
RMSEA = 132, CI .068-.198, p < .05 = .021
CFI = .956 SRMR = .040
CFI and SRMR suggest a good model fit. Chi-square and RMSEA do not. No mod. indices suggested or employed.

Advocate Reports at 4 Months

A4 Play

MI suggested A4E5 WITH A4E3 (same MI for the A4 Play model)

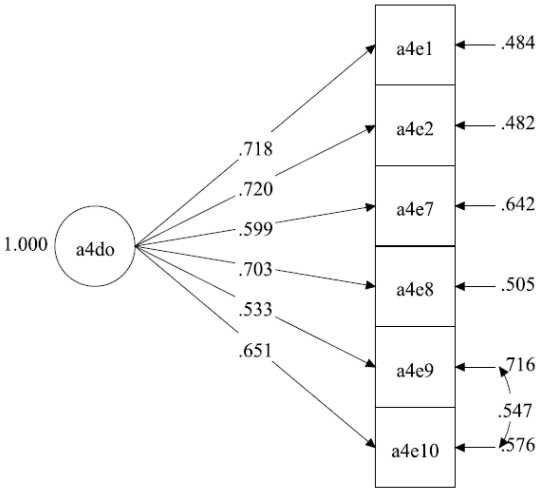
Chi square = .275, df 1, p = .5997, Chi square/df .275
RMSEA .000, CI .000-.209, p < .05 = .644
CFI = 1.00 SRMR = .009



Fit indices indicate a good model fit. No mod indices suggested or employed.

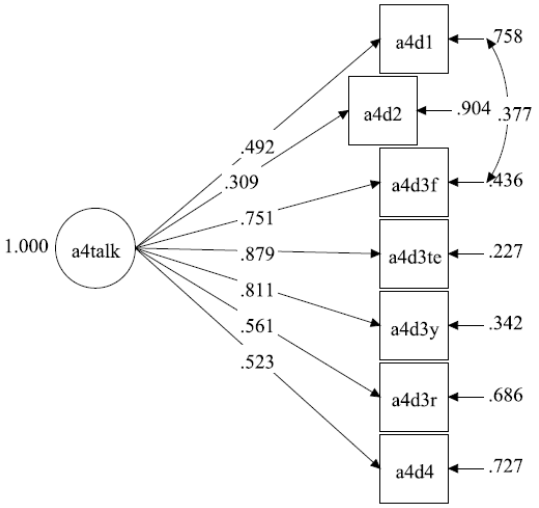
A4 Doing in the Community

MI suggested A4E9 WITH A4E10
20141103 A4Do 9w10
Chi square = 5.173, df 8, p = .7389, chi square/df = .647
RMSEA = .000, CI .000-.083, p < .05 = .853
CFI = 1.00 SRMR = .024



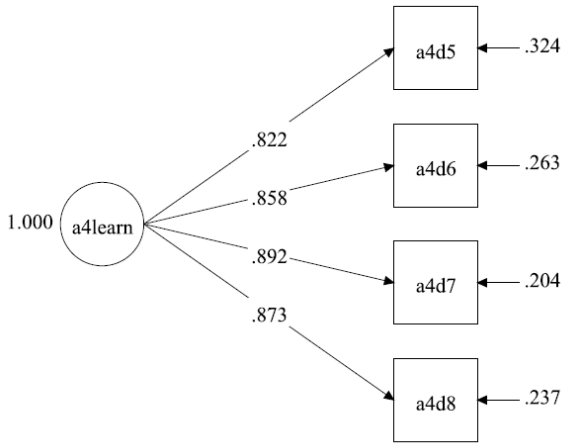
A4 Talk: Casual Conversation

Initially poor fit indices. Modification Indices suggested covarying A4D3F WITH A4D1
Revised:
Chi square = 26.32, df 13, p = .0154,
chi square/df = 2.025
RMSEA = .099, CI .042-.154, p < .05 = .071
CFI = .946 SRMR = .067
Chi square/df, CFI and SRMR suggest adequate model fit. No further mod. indices suggested or employed.



A4 Problem-focused Discussions

Chi square 6.687, df 2, p = .0353, chi square/df = 3.3435
RMSEA = .150, CI .034-.283, p < .05 = .069
CFI .985 SRMR = .017
CFI and SRMR suggest a good model fit. No mod. indices suggested or employed.



APPENDIX D
Items in Scales Created or Used for Exploratory Advocacy Analyses

Scale Name and Response Prompt	Items in Youth and Advocate Surveys
DuBois Advocacy Scale "How often does your Advocate do the following things **when you are together?"**	**F 1 … says or does something to help me with my feelings**
	F 2 …gives me advice or ideas about how to do something
	F 3 Shows me how to do something
	F 4 Helps me practice something
	F 5 Shows or tells me things about his or her own life
	F 6 Helps me think about myself or the world in a different way
	F 7 Helps me figure out what is really important in my life
	F 8 Helps me by talking with other people who are important in my life
	F 9 Makes sure I have the things I need to be successful
	F 10 Makes sure I have the chance to participate in activities that are good for me
	F 11 Accompanies me to visit his/her contacts or friends who might help me get a job, get into school, enroll in a training program, learn a new skill, etc.
	F 12 Help me sign up for an opportunity or program
Blue Ribbon Advocacy Scale "How often does your Advocate do any of the following things when you are not together?"	G 1 Talks with my family about how I am doing in school.
	G 2 (youth only) Goes to parent-teacher conferences or meetings
	G 3 Talks with my teacher/teacher aide
	G 4 Talks with my therapist/counselor.
	G 5 Communicates with people to help youth get a job, get into school, a training program
	G 6 Tells other people good things about me.
	G 6 (Advocate only) Tell other people (like teachers, boss) when you think the youth needs extra help from them
Adapted from Zand's (2009) *Mentor-youth Alliance Scale* "How true is each statement about how you feel about your Advocate?"	H 1 My Advocate cares about me
	H 2 My Advocate is happy when good things happen to me
	H 3 I would feel sad if something bad happened to my Advocate
	H 4 I try to follow my Advocate's advice
	H 5 I feel comfortable with my Advocate
	H 6 I enjoy talking with my Advocate
	H 7 I look forward to the time I spend with my Advocate

	H 8 I trust my Advocate
	H 9 My relationship with my Advocate is important to me
	H 10 I feel my Advocate cares about me, even when I do things s/he does not approve of
YAP Factor 1—Self Management Focus (Relational, individual) "How often do you and your Advocate individually…"	J 1 …work on life skills such, as healthy decision-making, dealing with stress and avoiding risky behaviors?
	J 5 …work to help me manage anger more appropriately?
	J 8 …try to help me avoid using drugs or alcohol?
YAP Factor 2—Vocational Development focus (Goal, individual) "How often do you and your Advocate individually…"	J 2 … work on job searchers, job applications, resumes or interviews?
	J 4 …. Work on independent living skills (budgeting, time management or use of public transportation
	J 6 … work to help me find a job (either a supported work job or regular paying job
	J 9 … discuss appropriate work place dress and behavior.
YAP Factor 3—Community Advocacy focus (relational, individual) "How often do you and your Advocate individually…"	J 3 … go meet an adult in the community who could be a mentor, role model or friend
	J 7 … meet with another youth in the community who could be a role model or friend
	J 10 … help me enrolling an arts, sports or recreation club, program or center in the community
YAP Factor 4—Focus with Family on Family Problem	Get help for a medical problem?
	Get help for a drug or alcohol problem?
	Get help for a mental health problem?
	Connect with a social organization (e.g., church, woman's cooperative, a social club) to address social isolation needs? Enroll in an educational or training program?
	Get emergency food or clothing or shelter?
	Find a job?
Factor 5—Focus with Family on Youth Problem	The youth's behavior at home?
	The youth's progress in YAP?
	The youth's legal situation.
	The youth's behavior, attendance or grades at school?
	The youth's behavior at home?

APPENDIX E.

YAP Advocacy Factors Correlates with Public Private Ventures Study Dataset

Correlations between YAP Advocacy Factors and Advocate Characteristics From Dataset Data Originally Collected for Public Private Ventures

		Self-Management (Rel) Focus	Vocational Development (Goal) Focus	Community Advocacy (Rel) Focus	Family Member Problem (Goal) Focus	Focus Family on Youth Problem
What is your age?	r	.059	.064	.075	**.112**	.109
	p	.358	.313	.241	.079	.088
	N	246	248	246	247	246
Are you female (1) or male (0)	r	-.035	.032	.020	.091	.015
	p	.588	.612	.752	.154	.810
	N	246	248	246	247	246
What is your race: Latino, Hispanic	r	-.073	-.031	-.054	-.052	.058
	p	.250	.630	.394	.416	.364
	N	249	251	249	250	249
What is your race or ethnicity: African American, Black	r	-.022	-.013	-.002	-.054	-.019
	p	.734	.834	.975	.398	.760
	N	249	251	249	250	249
What is your race or ethnicity: Caucasian, White	r	.044	-.026	-.037	-.035	-.034
	p	.488	.679	.563	.579	.596
	N	249	251	249	250	249
Can you speak any languages other than English? (Yes = 1)	r	-.003	-.040	-.037	-.122	-.027
	p	.958	.527	.561	.055	.671
	N	247	249	247	248	247

Factor Analyses, Item Loadings, and Intercorrelations of Items

	Factor			
	1: Problem focus	2: Help develop as a worker	3: Help foster community connections	4: Help Promote self dev-elopment
The youth's behavior, attendance or grades at school?	**.872**	.303	.417	.470
The youth's behavior at home?	**.838**	.317	.438	.474
The youth's progress in YAP?	**.690**	.371	.297	.490
The youth's legal situation.	**.580**	.308	.389	.435
Helped with job searches, job applications, resumes or interviews?	.297	**.957**	.435	.432
Helped find a job	.320	**.668**	.382	.367
Talked to about appropriate work place dress and behavior?	.365	**.639**	.412	.536
Introduced to another youth in the community who could be a role model or friend to the youth?	.405	.403	.814	.461
Introduced to an adult in the community who could be a mentor, role model or friend to the youth?	.320	.350	**.738**	.397
Helped enroll in an arts, sports or recreation club, program or center in the community?	.344	.414	**.609**	.380
Talked with the youth's schoolteacher or counselor about his/her academic progress or behavior?	.348	.297	**.448**	.403
Worked with individually to help him/her manage anger more appropriately?	.432	.334	.379	**.767**
Worked with individually to help him/her avoid using drugs or alcohol?	.446	.390	.412	**.714**
Worked with individually on independent living skills such as budgeting, time management or using public transportation?	.435	.577	.476	**.615**
Worked with individually on life skills such, as healthy decision-making, dealing with stress and avoiding risky behaviors?	.380	.413	.459	**.606**
Assisted with homework assignments or schoolwork?	.347	.361	.426	**.442**

APPENDIX F
Correlations Between YAP Advocacy Factors and Goals in the Advocates ISP's.

Some of the Advocate IEP goals were related to the YAP advocacy approaches.

		Self-Management (Rel) Focus	Vocational Development (Goal) Focus	Community Advocacy (Rel) Focus	Family Member Problem (Goal) Focus	Focus Family on Youth Problem
Self-Management (Rel) Focus	Pearson Correlation	1	.486**	.448**	.160*	.491**
	Sig. (2-tailed)		.000	.000	.011	.000
	N	249	249	249	248	247
Vocational Development (Goal) Focus	Pearson Correlation	.486**	1	.466**	.293**	.416**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	249	251	249	250	249
Community Advocacy (Rel) Focus	Pearson Correlation	.448**	.466**	1	.347**	.412**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	249	249	249	248	247
Family Member Problem (Goal) Focus	Pearson Correlation	.160*	.293**	.347**	1	.221**
	Sig. (2-tailed)	.011	.000	.000		.000
	N	248	250	248	250	249
Focus Family on Youth Problem	Pearson Correlation	.491**	.416**	.412**	.221**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	247	249	247	249	249
Plan my weekly activities with the youth.	Pearson Correlation	.117	.155*	.143*	.090	.068
	Sig. (2-tailed)	.069	.016	.027	.166	.297
	N	240	241	240	240	240
Plan how to address the youth' s educational needs.	Pearson Correlation	.173**	.218**	.231**	.091	.112
	Sig. (2-tailed)	.007	.001	.000	.161	.083
	N	240	241	240	240	240
Plan how to address the youth' s social and emotional needs.	Pearson Correlation	.184**	.205**	.204**	.097	.065
	Sig. (2-tailed)	.004	.001	.001	.135	.318
	N	240	241	240	240	240
Learn about the youth' s interests.	Pearson Correlation	.080	.164*	.163*	.062	.035
	Sig. (2-tailed)	.218	.011	.012	.344	.591
	N	238	239	238	238	238
Learn about the youth' s strengths.	Pearson Correlation	.107	.155*	.225**	.112	.037
	Sig. (2-tailed)	.097	.016	.000	.083	.568
	N	240	241	240	240	240
Plan what community programs and resources to link the youth with.	Pearson Correlation	.180**	.216**	.232**	.079	.088
	Sig. (2-tailed)	.006	.001	.000	.228	.176
	N	236	237	236	236	236
Understand the youth' s relationships with his or her family.	Pearson Correlation	.199**	.224**	.256**	.089	.054
	Sig. (2-tailed)	.002	.001	.000	.171	.409
	N	237	238	237	237	237
Keep track of youth' s progress in the program.	Pearson Correlation	.130*	.226**	.236**	.063	-.014
	Sig. (2-tailed)	.045	.000	.000	.332	.823
	N	240	241	240	240	240

APPENDIX G
Construction of and Factorial Validity Evidence for the two New Advocacy Scales

The DuBois' Derived D-AIR Scale

In response to the finding, in his latest meta-analytic review of the effectiveness of youth mentoring (DuBois et al., 2011), that one moderator of effectiveness was whether or not programs that emphasized advocacy and teaching in their mentoring model. Measured collectively, those programs with an advocacy and teaching focus were more effective. This not only prompted the research funding initiative that supported this study, but also led DuBois himself to start to explore advocacy.

One way he did so was to create a list of items, some of which directly assessed advocacy (items 8-12 below). As there was no prior scale construction or validation work done on this set of items (below), we collected self-reports on these items and then undertook a series of factor analyses to, we assumed, separate the different functions of mentors according to which were advocacy, teaching, or other approach.

However, our initial exploratory analyses did not reveal the items to reflect distinct phenomena but rather one general factor. All 12 of the items DuBois developed also loaded onto one factor when subjected to confirmatory factor analyses in separate analyses of both two and four month (discharge data). It therefore seemed what DuBois had created was a measure of advocacy, instruction, and reflection—a more comprehensive assessment of advocacy, teaching, and encouragement. Therefore, with permission (but not approval, per se) from David DuBois, we decided to include this scale as a multi-dimensional measure of advocacy within the context of instruction and personal reflection. Although this does not provide "pure assessment" of interpersonal advocacy—note, by contrast, the Blue Ribbon Advocacy Scale is about advocacy outside the youth-adult relationship—the measure developed from DuBois items is not unlike the moderator in his meta-analysis which included both teaching and advocacy in the context of mentoring, which we feel this scale nicely captures.

1. says or does something to help me with my feelings
2. gives me advice or ideas about how to do something
3. shows me how to do something
4. helps me practice something
5. shows or tells me things about his/her own life
6. helps me think about myself or the world in a different way
7. helps me figure out what is really important to me
8. helps me by talking with other people who are important in my life
9. makes sure I have the things I need to be successful
10. makes sure I have a chance to participate in activities that are good for me
11. accompanies me to visit his/her contacts or friends who might help me get a job, get into school, enroll in a training program, learn a new skill, etc..
12. help me sing up for an opportunity or program

(In the analyses that follow, it is revealed that items 11 and 12 had high correlations with the Blue Ribbon scale factor and were subsequently omitted to create orthogonal measures.)

The Graig Meyers Derived Blue Ribbon Advocacy Scale

Then we shifted our focus on testing a set of items that a leader in the mentoring field on the use of advocacy in youth mentoring, Graig Meyers who directs the Blue Ribbon Youth Mentoring Program proposed. The items he proposed as examples or key indicators of advocacy were not meant to reflect all of the dimensions of advocacy in youth mentoring, nor did he intend for these to become prompts or items in a survey. He had simply responded on DuBois mentoring listserv to a members request for information about what "advocacy" was. The question was referring to the DuBois et al. (2013) meta-analysis finding that programs emphasizing advocacy and teaching had larger effects.

With his permission, using his examples (see email on next page), we made minor changes to his prompts in the listserv posting, and subjected them to exploratory and confirmatory factor analysis with data collected for this study.

These six items, based on the items in his original posting which can be found in the appendix, were found to load on one statistical factor in exploratory analyses that provided good model fit indices when then subjected to confirmatory factor analyses.

1. talks with my family about how I am doing in school
2. goes to parent-teacher conferences or meetings
3. talks with my teacher/teacher aide
4. talks with my therapist/counselor
5. communicates with people to help youth get a job, get into school, a training program or other program
6. tells other people good things about me

Constructing Orthogonal Scales for the Present Study Purposes

The final step we took before any analyses that would compare these blended and pure assessments of advocacy was to subject them to both exploratory and confirmatory factor analyses to ensure that they actually reflected distinct approaches and that items did not want to load on both, in which case it would be hard to make interpretations about what each factor or subscale was measuring. When we did these analyses, we decided that two variables on each of the original scales should be removed, at least for the present study and namely for the purpose of making independent comparisons of the relative utility of each approach (or scale) in predicting changes observed following program participation.

When these two sets of items were subjected to these exploratory and confirmatory factor analyses, we learned that items 1 and 6 had problematically large correlations with the DuBois advocacy scale factor—they were redundant with what was being measured by DuBois items. Therefore, those two items were subsequently omitted. Subjected again to confirmatory factor analysis, the resulting 4-item Blue Ribbon Advocacy (BRA) scale had equally good fit indices, suggesting it too represented one construct (presumably advocacy) and which would be more orthogonal to (independent of) DuBois' AIR scale.

Still, there remained two of the AIR items that wanted to load on the Blue Ribbon scale. The last two items (11 and 12) were two of five of the DuBois items that DuBois had viewed as specifically about advocacy. These seemed to represent the youth's perceptions of things the Advocate does—how the Advocate Advocates for the youth—when the Advocate is with the youth to make connections across the youth's social ecologies of school, work and other programs. Conversely, the Blue Ribbon scale is more about when the Advocate does this sort of connecting and communicating when not with the youth. Nevertheless, to create orthogonal scales, items 11 and 12 were removed from the DuBois AIR scale, making it more blended with only three advocacy items (9 and 10), 4 about instruction and guidance (2, 3, and 4), and 3 items encouraging youth self-reflection. This teaching might better be called guidance in that it includes attention to the unique needs and circumstances of the youth—more like the Teacher+ of the TEAM framework. One of the self-reflection items identifies the Advocate modeling self-reflection by personally sharing their own experience (item 5) perhaps partly as a way of teaching but that is personal and directly related to the youth, such that it also reflects the collaboration indicative of a Counselor+ or Friend+ in the TEAM Framework.

Similarly, the Blue Ribbon Scale had just 4 (of 6 original) items that did not double-load (or duplicate items in the DuBois scale) and were not redundant with the AIR scale. Items 1 and 6 were more interpersonal, and item 6 could have happened in the presence of the youth (or else how would the youth know about it). Both were unlike the other four in that they did not clearly represent the conventional, future oriented, goal directed purpose in the TEAM framework that aimed to help improve the youth's success (academic, vocational) and future. The four remaining items on this scale are all about trying to effect change in the youth's world and trying to bridge the worlds of work, school, family, and mental health staff.

With the two modified scale capturing sufficiently distinct elements of advocacy, the analyses reported in the following section were undertaken. In their interpretation, we viewed the Blue Ribbon Advocacy scale as specifically about advocacy as outreach and fostering collaborations across contexts outside the youth-adult relationship, and the DuBois' Interpersonal AIR scale representing the interpersonal, psychosocial, skill-building, and "got your back" or "by your side" advocacy effort that may be viewed as a way to support the youth's confidence, skill levels, and self understanding in order to improve the youth's capacity to Advocate for him or herself presently and in the future independently.

Background: Source of the Blue Ribbon Advocacy scale. This scale was created using the following information from email summary from Graig Meyer (with permission)

"Most students who enter our program are unfamiliar with the idea of advocacy. We describe [advocacy] to them as

"when someone helps you even when they're not with you at the time."

It can take a little bit of work for the kids to get their heads around that, but here are some common examples we'd share with them:

-Your mentor will talk with your family about how you're doing in school.
-Your mentor will go with your family to parent-teacher conferences.
-Your mentor may talk with your teachers about how you're doing.
-Your mentor may research opportunities that would benefit you.
-Your mentor may help you sign up for an opportunity.
-Your mentor may tell other people good things about you.
-Your mentor may tell other people (like teachers) when they think you need extra help from them

Practitioners may want to consider how to train mentors to do those things and other advocacy activities. I hope this helps!
Graig Meyer, Blue Ribbon Mentor-Advocate, Chapel Hill-Carrboro City Schools
www.BlueRibbonMentors.org

In order to better differentiate the two Advocacy scales, items were subjected to an exploratory factor analysis. This revealed a possible third factor, but one with few items and that which required cross-loadings for decent model fit.

Rotated Component Matrix

	Component		
	1	2	3
Y2 DuBois Adv 1	.804		.136
Y2 DuBois Adv 2	.832		.226
Y2 DuBois Adv 3	.841		
Y2 DuBois Adv 4	.779		
Y2 DuBois Adv 5	.641	.288	-.201
Y2 DuBois Adv 6	.829	.187	
Y2 DuBois Adv 7	.803		.190
Y2 DuBois Adv 8	.786	.174	.189
Y2 DuBois Adv 9	.772		.323
Y2 DuBois Adv 10	.711		.285
Y2 DuBois Adv 11	.558	.393	
Y2 DuBois Adv 12	.633	.357	
Y2 Blue Ribbon Adv 1	.227	.300	.676
Y2 Blue Ribbon Adv 2		.867	.122
Y2 Blue Ribbon Adv 3		.879	
Y2 Blue Ribbon Adv 4	.128	.780	.101
Y2 Blue Ribbon Adv 5	.208	.446	.391
Y2 Blue Ribbon Adv 6	.142		.877

Once items with significant cross-over (needing covariances across scale factors to achieve good model fit—viz. DuBois 11 and 12, Blue Ribbon 1) and those pulling to create a third factor (Blue Ribbon 1 and 6), two factors emerged, one for each of the original set of items. Subsequently we fit a confirmatory factor analyses (CFA) of this two-factor model of advocacy.

Rotated Component Matrix

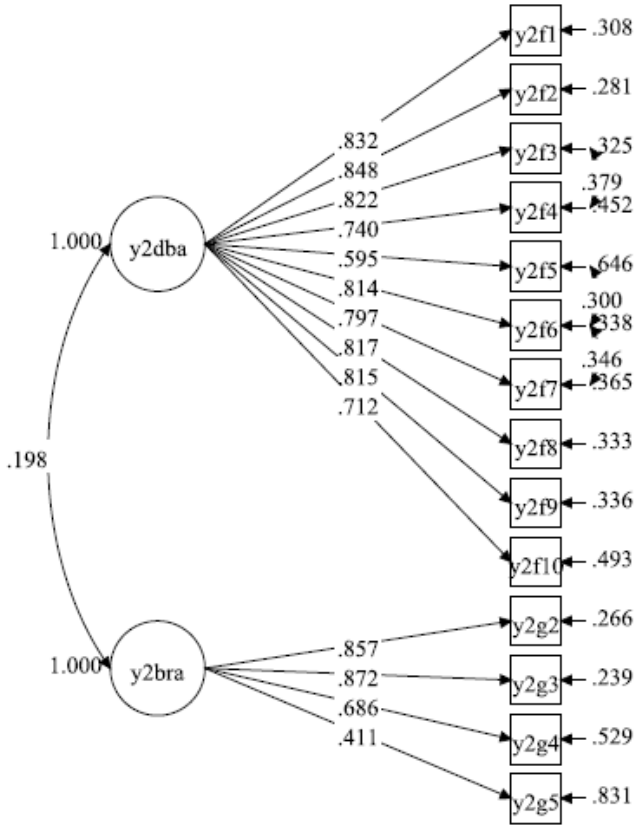
	Component	
	1	2
Y2 DuBois Adv 1	.819	
Y2 DuBois Adv 2	.861	
Y2 DuBois Adv 3	.847	
Y2 DuBois Adv 4	.769	
Y2 DuBois Adv 5	.589	.249
Y2 DuBois Adv 6	.827	.205
Y2 DuBois Adv 7	.829	.107
Y2 DuBois Adv 8	.804	.216
Y2 DuBois Adv 9	.820	.129
Y2 DuBois Adv 10	.748	
Y2 Blue Ribbon Adv 2		.882
Y2 Blue Ribbon Adv 3		.867
Y2 Blue Ribbon Adv 4	.121	.803
Y2 Blue Ribbon Adv 5	.264	.553

Pictured at right, this model, for the data collected at two months from the youth, fit the data reasonable well, with the SRMR and the Chi-Square to degrees of freedom ratio indices most suggestive of a good model fit.

Chi square = 156.527, df = 73, p = .000, chi square/df = 2.144
RMSEA = .095, CI .074-.115, $p <$ .05 = .000 CFI = .929 SRMR = .08

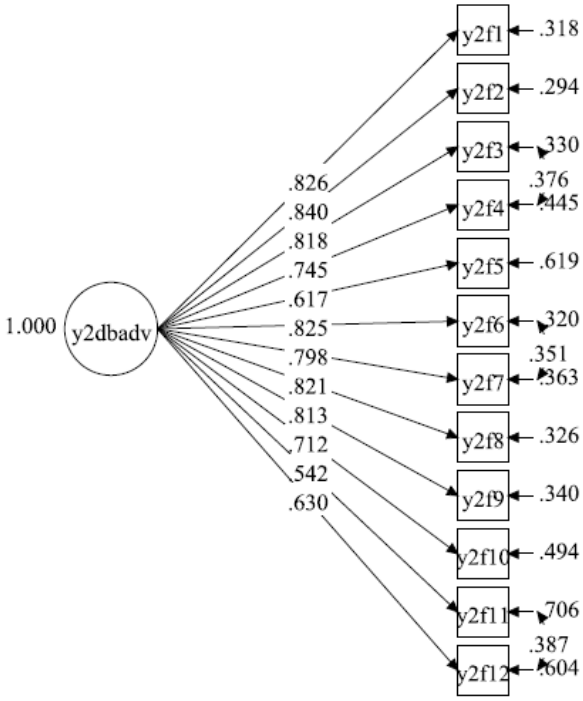
Note: This DuBois factor model allowed within-factor measure cross-loadings for items 3 with 4, 5 with 6, and 6 with 7.

Identifying the source of this covariance in subsequent work may reveal items which might be excluded from the scale.



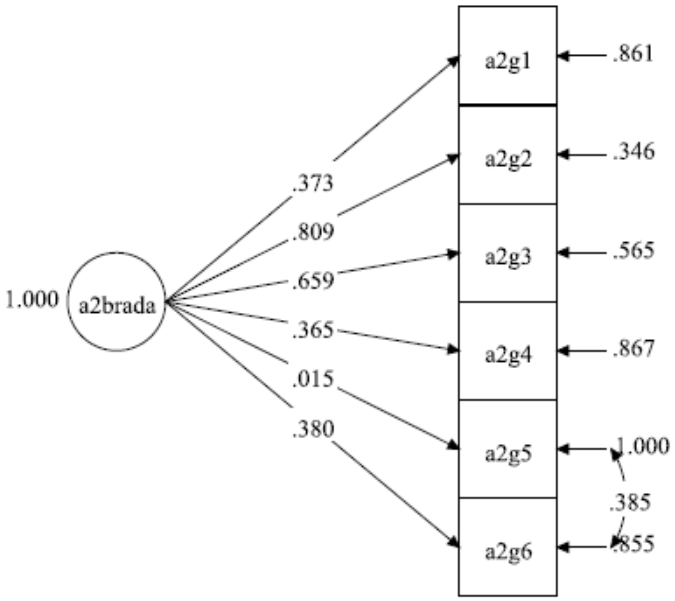
Differentiating the two Advocacy scales for Use in the Study
Separate scale factor Alternative Model Fit Indices and Analyses

DuBois Advocacy Youth-Reported Full Scale CFA Model Fit at 2 Months



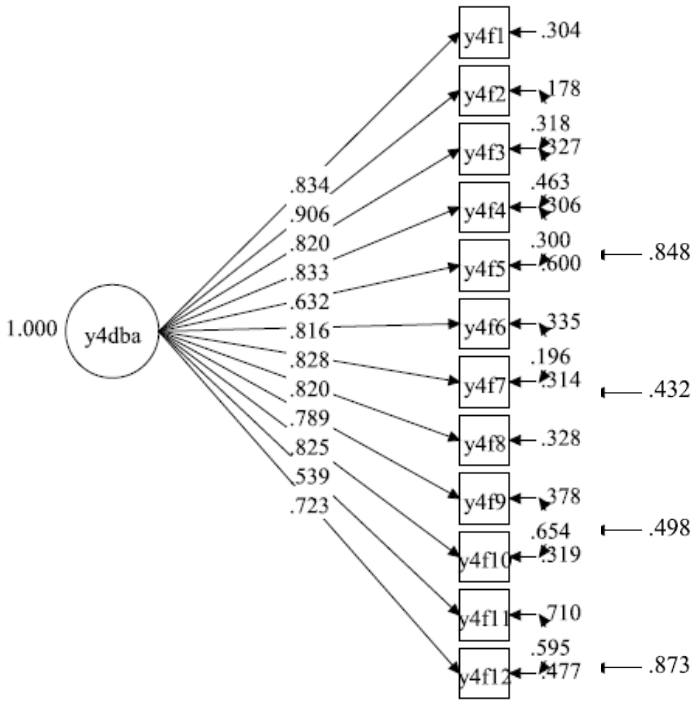
Chi-square 109.298, df 51, p = 0000, chi square/df = 2.143. RMSEA = .095, CI .070-.119, $p < .05$ = .002; CFI = .947, SRMR = .040. Chi square/df, CFI and SRMR suggest adequate model fit using the original items.

The Blue Ribbon Scale Full Scale CFA Model Fit at 2 Months
The Blue Ribbon Scale Fit Indices: Original 6-item and Final 4-item Model Used in Analyses



Chi-square = 19.448, df = 8, p = .0126, chi square/df = 2.431. RMSEA = .103, CI .45-.162, $p < .05$ = .064. CFI = .897. SRMR = .065.

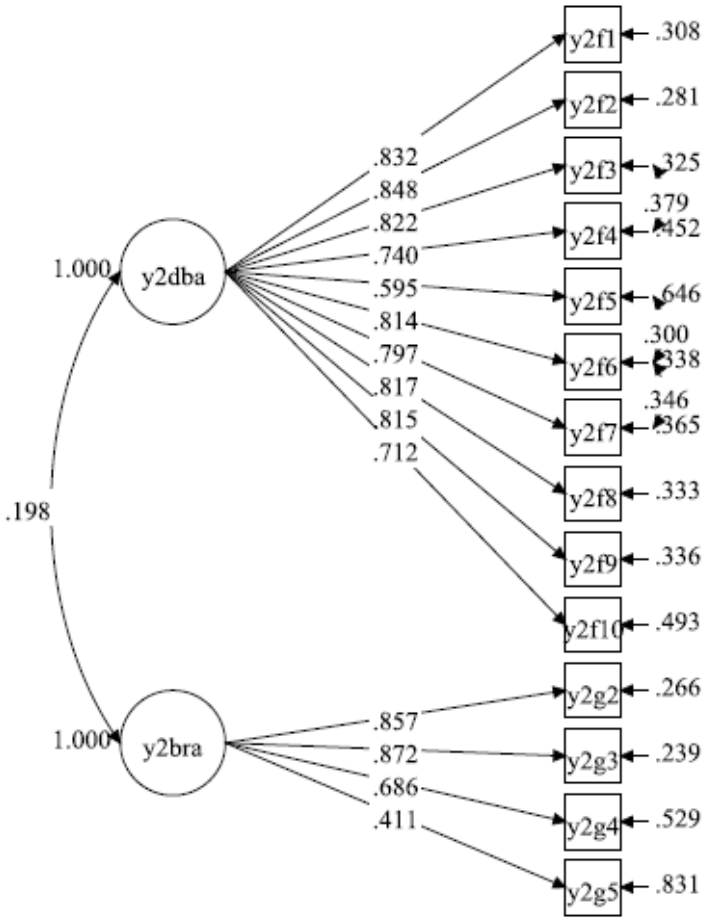
The SRMR is the only fit index that suggests a good model fit.



Chi square = .433, df 2, p = .8055
RMSEA = .000, CI .000-.105, *p* < .05 = .857. CFI = 1.0
SRMR = .012

All of the fit indices suggest this is a good fitting model. Fit indices for a two-factor CFA of the two measures of advocacy, each with two cross-loading items removed, suggesting separate constructs are captured

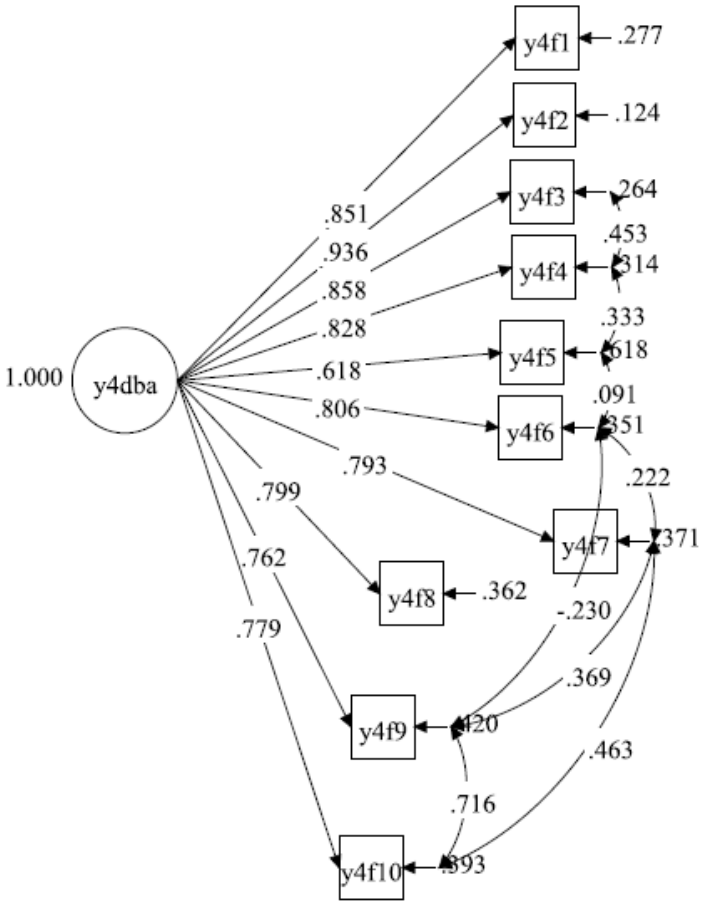
Chi-square = 156.527, df = 73, p = .000, chi square/df = 2.144. RMSEA = .095, CI .074-.115, *p* < .05 = .000, CFI = .929
SRMR = .08



Two CFA Factor Models of the Youth-Reported DuBois Advocacy with data at four months

Full scale CFA with data at four months:

Chi square = 99.631, df 48, *p* = .000, chi square/df = 2.0756
RMSEA = .120, CI .074-.131, *p* < .05 = .002. CFI = .955.
SRMR = .040

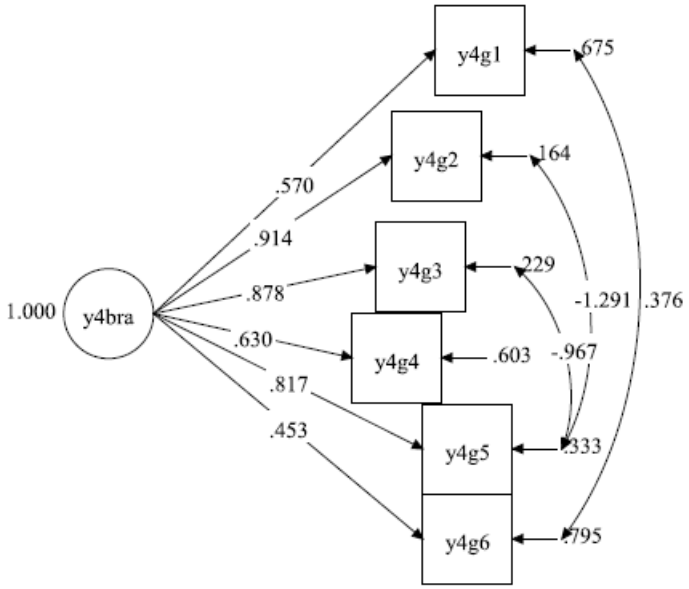


Scale CFA with data at four months, omitting the two items that cross-load onto the Blue Ribbon Advocacy Scale:

Chi square = 51.255, df = 27, p = .0032, Chi square/df = 1.8983 RMSEA = .093, CI .053-.132, $p <$ .05 = .040 CFI = .97 SRMR = .032
Model fit is good. At the same time, there are the three covariance relationships between the three pairs of items identified in the two month data that must be allowed in the model to achieve this model fit.

Factor Models of the Youth-Reported Blue Ribbon Advocacy Scale at four Months

Full scale CFA with data at four months for youth-reported Blue Ribbon Advocacy:

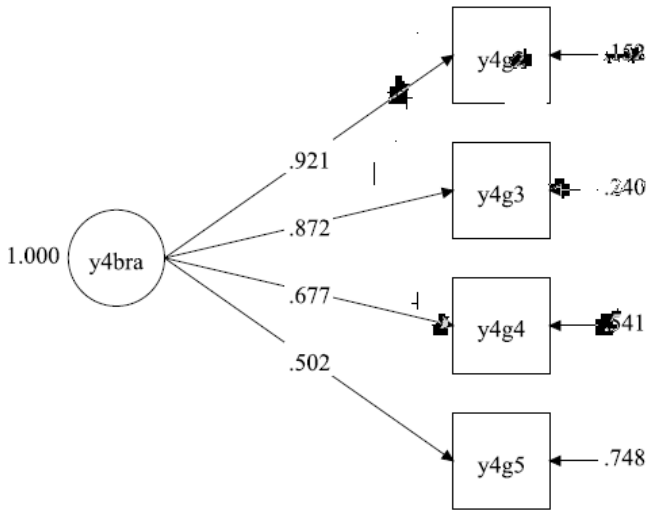


Chi square = 18.660, df 6, p =.004. Chi square/df = 3.11 RMSEA = .145, CI .073-.221, $p < .05$ = .019 CFI = .953 SRMR = .058

Chi square/df and RMSEA suggest poor model fit, while the CFI and SRMR suggest good model fit.

The required cross-loadings, however, are problematic.

CFA with no cross-loading items with data at four months for Blue Ribbon Advocacy:



Chi square = 5.385, df 2, p = .0677, chi square/df = 2.6925 RMSEA = .129, CI .000-.267, $p < .05$ = .116 CFI = .981 SRMR = .036

APPENDIX H

Hemingway Measure of Adolescent Connectedness Items (see Karcher & Sass, 2010)

Items 1-30 assess connectedness to friends, family, school, teachers, and the future. Each scale has 6 items. Connectedness to friends, for example, is the average of items 1, 6, 11, 16, 21 and 27. One reverse worded item is in each subscale.

These questions ask how you feel about things like school, other kids your age, and yourself. For each sentence, decide how true the sentence is for you. Then circle one number that fits best. "How TRUE about you is each sentence?

	Not at all (1); not really (2); sort of (3); true (4); very true (5)				
(1) Spending time with friends is not so important to me.	1	2	3	4	5
(2) My family has fun together.	1	2	3	4	5
(3) I work hard at school (or other educational program like GED).	1	2	3	4	5
(4) I care what my teachers think of me.	1	2	3	4	5
(5) I will have a good future.	1	2	3	4	5
(6) I have friends I'm really close to and trust completely.	1	2	3	4	5
(7) It is important that my parents trust me.	1	2	3	4	5
(8) I enjoy being at school (or the educational program I attend).	1	2	3	4	5
(9) I do not get along with some of my teachers.	1	2	3	4	5
(10) Doing well in school will help me in the future.	1	2	3	4	5
(11) Spending time with my friends is a big part of my life.	1	2	3	4	5
(12) I enjoy spending time with my parents.	1	2	3	4	5
(13) I get bored in school a lot.	1	2	3	4	5
(14) I want to be respected by my teachers.	1	2	3	4	5
(15) I do things outside of school to prepare for my future.	1	2	3	4	5
(16) My friends and I talk openly with each other about personal things.	1	2	3	4	5
(17) My parents and I disagree about many things.	1	2	3	4	5
(18) I do well in school (or the educational program I attend now).	1	2	3	4	5
(19) I try to get along with my teachers.	1	2	3	4	5
(20) I do lots of things in school to prepare for my future.	1	2	3	4	5
(21) I spend as much time as I can with my friends.	1	2	3	4	5
(22) My parents and I get along well.	1	2	3	4	5
(23) I feel good about myself when I am at school.	1	2	3	4	5
(24) I always try hard to earn my teachers' trust.	1	2	3	4	5
(25) I think about my future often.	1	2	3	4	5
(26) I usually like my teachers.	1	2	3	4	5
(27) My friends and I spend a lot of time talking about things.	1	2	3	4	5
(28) I care about my parents very much.	1	2	3	4	5
(29) What I do now will not affect my future.	1	2	3	4	5
(30) Doing well in school is important to me.	1	2	3	4	5

Select References

Broussard, C. A., Mosley-Howard, S., & Roychoudhury, A. (2006). Using youth Advocates for mentoring at-risk students in urban settings. *Children & Schools, 28*(2), 122-127.

Davidson, W. S., II. (1976). *The Diversion of Juvenile Delinquents: An Examination of the Processes and Relative Efficacy of Child Advocacy and Behavioral Contracting.* Ph.D., University of Illinois at Urbana-Champaign, Illinois.

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology, 30*, 157-197.

DuBois, D. L., Neville, H. A., Parra, G. R., & Pugh-Lilly, A. O. (2002). Testing a new model of mentoring. *New Directions for Youth Development, 93*(Spring), 21-57.

DuBois, D. L., Portillo, N., Rhodes, J. E., Silverthorn, N., & Valentine, J. C. (2011). How effective are mentoring programs for youth? A systematic assessment of the evidence. *Psychological Science in the Public Interest, 12,* 57-91.

Gibson, C.L., Miller, J. M., Jennings, W. G., Swatt, M., & Gover, A. R. (2009). Using propensity score matching to understand the relationship between gang membership and violent victimization: A research note. *Justice Quarterly, 26,* 625-643.

Grossman, J. B., & Tierney, J. P. (1998). Does mentoring work? An impact study of the Big Brothers Big Sisters program. *Evaluation Review, 22,* 3, 403-426.

Halpern, R. (2005). Instrumental relationships: A potential relational model for inner-city youth programs. *Journal of Community Psychology, 33*(1), 11-20. doi: 10.1002/jcop.20032

Hamilton, S. F., & Hamilton, M. A. (1992). Mentoring programs: Promise and paradox. *Phi Delta Kappan*(73), 546-550.

Hamilton, M. A., & Hamilton, S. F. (2005). Work and service-learning. In D. L. DuBois & M. J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 348-363). Thousand Oaks, CA: Sage Publications Ltd.

Herrera, C., Grossman, J. B., Kauh, T. J., Feldman, A. F., McMaken, J., & Jucovy, L. Z. (2007). *Big Brothers Big Sisters school-based mentoring impact study*. Philadelphia: Public/ Private Ventures.

Jessor, R., & Jessor, S. L. (1977). *Problem behavior and psychological development: A longitudinal study of youth*. New York: Academic Press.

Johnson, W. B. (2008). Are advocacy, mutuality, and evaluation incompatible mentoring functions? *Mentoring & Tutoring: Partnership in Learning.*

Special Issue: The ethics of mentorship, 16(1), 31-44.

Jolliffe, D., & Farington, D. P. (2007). A rapid evidence assessment of the impact of mentoring on re-offending: A summary. Cambridge University: Home Office Online Report 11/07. http://www.crimereduction.gov.uk/workingoffenders/workingoffenders069.htm

Karcher, M. J. (2008). The Study of Mentoring in the Learning Environment (SMILE): A randomized evaluation of the effectiveness of school-based mentoring. *Prevention Science, 9*(2), 99-113. doi: 10.1007/s11121-008-0083-z

Karcher, M. J., Davidson, A., Rhodes, J. E., Herrera, C. (2010). Pygmalion in the program: The role of teenage mentors' attitudes in shaping their mentees' outcomes. *Applied Developmental Science, 14*(4), 212-227.

Karcher, M. J. & Hansen, K. (2014). Mentoring activities and interactions. In D. L. DuBois, & M. J. Karcher (Eds.), *Handbook of youth mentoring* (2nd Edition; pp. 63-72). Thousand Oaks, CA: Sage Publications.

Karcher, M. J., Herrera, C., & Hansen, K. (2010). "I dunno, what do you wanna do?": Testing a framework to guide mentor training and activity selection. *New Directions for Youth Development, 126*, 51-69. doi: 10.1002/yd.349.

Karcher, M. J. & Johnson, D. J (2014, Nov. 20). An Evaluation of Advocacy-Based Mentoring as a Treatment Intervention for Chronic Delinquency: Initial Findings. Paper Symposium entitled Evaluating Advocacy-based Mentoring Models to Prevent and Reduce Delinquency, *American Society of Criminology*, San Francisco, CA.

Karcher, M. J., & Nakkula, M. J. (2010). Youth mentoring with a balanced focus, shared purpose, and collaborative interactions. *New Directions in Youth Development, 126* (Summer), 13-32

Karcher, M. J. & Sass, D. (2010). A multicultural assessment of adolescent connectedness: Testing measurement invariance across gender and ethnicity. *Journal of Counseling Psychology, 57*(3) 274-289.

Keller, T. E., & Pryce, J. M. (2012). Different roles and different results: How activity orientations correspond to relationship quality and student outcomes in school-based mentoring. *The Journal of Primary Prevention, 33*(1), 47-64. doi: 10.1037/0003-066x.55.1.122

Liang, B., Spencer, R. e., Brogan, D., & Corral, M. (2008). Mentoring relationships from early adolescence through emerging adulthood: A qualitative analysis. *Journal of Vocational Behavior, 72*(2), 168-182. doi: 10.1111/b.9781405133739.2007.00024.x

McDonald S., Erickson L. D., Johnson M. K., & Elder G. H. (2007). Informal mentoring and young adult employment. *Social Science Research, 36,*

1328–1347.

McPartland, J. M., & Nettles, S. M. (1991). Using community adults as advocates or mentors for at-risk middle school students: A two-year evaluation of Project RAISE. *American Journal of Education, 99,* 568-586.

Morrow, K. V., & Styles, M. B. (1995). *Building relationships with youth in program settings: A study of Big Brothers/Big Sisters*. Philadelphia, PA: Public/Private Ventures.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24,* 315-340.

Nakkula, M. J., & Harris, J. T. (2010). Beyond the dichotomy of work and fun: Measuring the thorough interrelatedness of structure and quality in youth mentoring relationships. *New Directions for Youth Development, 2010*(126), 71-87. doi: 10.1002/yd.350

Parra, G. R., DuBois, D. L., Neville, H. A., Pugh-Lilly, A. O., & Pavinelli, N. (2002). Mentoring relationships for youth: Investigation of a process-oriented model. *Journal of Community Psychology, 30*(4), 367-388.

Rhodes, J. E. (2002). *Stand by me: The risks and rewards of youth mentoring.* Cambridge: Harvard University Press.

Tolan, P., Henry, D., Schoeny, M., & Bass, A. (2008). Mentoring interventions to affect juvenile delinquency and associated problems. *Campbell Systematic Reviews 16*: 1-112.