Choice Research Associates
What gets measured gets done.

Good Shepherd Mediation Program

Juvenile Offender Diversion Programs
Youth Delinquency and Violence Prevention Workshop

Final Report

September 2018

***

Shawn M. Flower, Ph.D.
Principal Researcher
and
Erin J. Farley, Ph.D.
Bailey Maryfield, M.S.

***

P.O. Box 322 ♦ Greenbelt, MD 20768-0322 ♦ Tel: 301-552-9567
www.choiceresearchassoc.com

Table of Contents

List of Tables

List of Figures

List of Appendices

Executive Summary

Choice Research Associates (CRA) was contracted to examine the characteristics and recidivism rates of 4,023 youth referred to the Good Shepherd Mediation Program (GSMP) from 2008 to 2017.

GSMP offers services to youth who participate in the Juvenile Offender Diversion Program (JODP). This program is available to first-time offenders between the ages of 11 and 17 who have been arrested for a misdemeanor (e.g., underage drinking, loitering, disorderly conduct, defiant trespassing, harassment, vandalism, criminal mischief, terroristic threats, possession of a weapon on school property, shoplifting, auto-related offense) or a simple felony (e.g., assault).

GSMP also offers services to youth who participate in the Youth Delinquency and Violence Prevention (YDVP) program. This program serves youth between the ages of 11 and 17 who have not been arrested but instead have been identified as at-risk of delinquency based on risky behavior that does not rise to the level of delinquency (e.g., truancy, peer associations, verbal or physical aggression). Among these participants, 69% were referred by a Youth Aid Panel (YAP) and 29% were court ordered.

Youth in both JODP and YDVP are required to attend a conflict resolution workshop. This workshop is based on restorative-justice principles and designed to reduce recidivism. These programs offer at-risk and delinquent youth the opportunity to learn productive ways to manage conflict, control their anger, and make better decisions. Youth who committed an auto theft or retail theft offense are required to participate in an auto theft or retail theft module during the conflict resolution workshop. These offense specific modules focus on the impact of these offenses on the youth, the victim, and the community. Youth who participated in GSMP completed a pre- and post-survey which assessed attitudes on physical violence, anger control, conflict resolution, understanding consequences of actions, helpfulness of the workshop.

This report explores demographics and outcomes of GSMP participants over a 10 year period, while also breaking out the demographic and outcomes for those who participated in the auto theft and retail theft modules.

Program participants were on average 15 years old, 66% were male, 75% were Black/African American, 8% White, and 17% identified as other race. On average, GSMP served on average 400 youth. Among these youth, 33% had been charged with a person offense (e.g., assault, robbery, harassment); 33% were charged with a property offense (auto theft, vandalism, theft, burglary); 14% had a weapons offense (possession of a gun, knife, or box cutter on school property); 12% were charged with a drug crime (possession, use, and sales of all types of drugs); 6% had public order offenses (including disorderly conduct, resisting arrest and terroristic threats); and 2% had "Other" type offenses including conspiracy and community referrals).

This report explores demographic and offense differences by type of program, and significant differences were found. For example, over 90% of youth in the auto theft module are male compared to 30% of those in the retail theft module. In addition, results show significant variation across the three offense groups by Ethnicity, with Hispanics more likely to be

associated with "other" offense. By race, 79% of those with person or weapon offenses were African American/Black and Caucasian/White youth are disproportionately represented among drug offenders at 14%. All demographic variables except Ethnicity (Hispanic) also significantly varied across the offense types. Those with a weapon offense were the youngest with a mean age of 14.61 years where those with a drug offense were older with a mean age of 16.14 years. There were gender differences as well – among youth charged with a drug offense, 86% were male and 14% were female. In comparison, 52% of those with weapon offenses were male and 48% were female. The analysis also reveals variation by the referral source. For example, 87% of youth associated with a drug offense were referred by YAP, while 15% of youth were Court Ordered while among those with person related offense 52% were referred by YAP and 46% were Court Ordered.

There are two primary outcomes examined in this report – 1) pre-to-post self-reported surveys and 2) whether participants were arrested after participating in GSMP. The surveys included 23 statements to which they could select from a 5-point Likert scale assessing their level of agreement (1 = strongly disagree to 5 = strongly agree). Overall, 85% (3,429) of participants had data for both the pre- and post-survey. Each individual survey item was examined to determine if there were significant differences comparing before and after the workshop.

Key findings included that participants were (on average) ***more*** likely to agree with the following statements:

- It is okay to walk away from a fight whether or not you think you would win.
- I know what things people do or say that trigger my anger.
- It's possible for me to think about the consequences of my behavior before I act on my feelings.
- I can choose different ways of reacting to someone when they make me mad.
- It's possible for people to see the same situation in different ways.
- I see how my actions have affected my relationships with family, friends, and others.

Overall these survey results indicate that participants gained knowledge on avoiding physical violence, anger control, conflict avoidance, thinking of consequences, understanding how actions affect others, and an understanding of and appreciation for participation in the program.

The relationship among the 23 survey questions was examined to determine the existence of underlying factors that can be used to create scales (referred to as latent constructs). While GSMP staff identified 7 themes, using correlation, factor and reliability analysis were conducted, 4 constructs (or scales) were constructed from the pre-test survey data. The 4 scales are:

- **Construct 1 - Actions/Consequences:** combines 7 survey questions focused on personal responsibility, understanding the seriousness of the offense, and the potential impact.
- **Construct 2 – Anger/Perspective:** contains 5 questions that relate to reacting and justifying feelings of anger, the utility of fighting, and value of listening.
- **Construct 3 – Program Expectations:** includes 4 questions centered on understanding why they are in the program and their openness to the experience.

- **Construct 4 – Agency/Potential:** involves 7 questions that touch on perspectives of control of feelings and actions – including ability to think about the consequences prior to taking action; and understanding that others have a different view of the same situation.

We explored the differences in the average pre-survey scores of each construct by demographic, program type, referral, and offense types. We provide examples of significant differences by:

- **Race:** White youth had higher pre-test average scores on constructs 1, 2 and 4;
- **Ethnicity:** Hispanic youth reported lower scores on Construct 4;
- **Gender:** female participants higher scores than male youth on Constructs 1 and 3;
- **Module or Workshop:** Participants in the retail theft module had higher scores on Constructs 1 and 3;
- **Referral Type:** those referred by YAP had higher scores than Court ordered youth on Constructs 1 and 3; and
- **Offense Type**: youth referred for a person related offense reported significantly lower mean scores than those with property, drug or weapon offenses on Constructs 1 and 4; and lower than property offenders on Construct 2.

Taken together these results indicate these diverse groups of youth (e.g., by race, gender, etc.) do not have the same "starting point" when they first come to the GSMP workshop. The next question is -- do these factors impact the level of change in survey outcomes? We examine this by conducting a regression analysis.

The 4 regression models include the same demographic and offense measures, informed by the analysis which found significant differences in the constructs. While the overall variation explained in the models is small (variables in the model only explaining from less than 1% to 2% of the outcomes), overall, across 3 of the 4 scales (C1, C2, C3), the key factors are the age of youth, gender, and type of offense (with more serious offenders showing the greatest impact on these outcomes). Specifically:

- **C1 –-the Actions/Consequences:** Significant predictors are age of youth, being male, white, and a person, drug, or weapon offender all significantly impact change from pre- to post-test. Older youth exhibit *more* change on this scale, as do those charged with a person, drug or weapon offense. Gender and race have a negative relationship with the scale – so being male and white youth have less change than female or non-white youth.
- **C2 – Anger/Perspective:** 4 predictor variables are significant – age of youth, male, person and weapon offender with males showing the strongest impact. Older youth, those charged with a person or weapon offense score have a positive impact on C2, while if you are male, you are *less* likely to report an improvement in perceptions on the C2 scale.
- **C3 – Program Expectations:** 6 variables are positively related to the outcome – age, person, drug, weapon, and public order offenders, and those referred by the court. Here person and weapon offenders have the strongest influence on change from on this scale.
- **Construct 4 – Agency/Potential:** There are 2 significant variables – Male participants have significantly less change on this scale than females, and person offenders have more change. However, the overall model was not statistically significant.

In addition to the Likert scale questions, participants had the opportunity to respond to six narrative questions in the post-survey. These questions asked youth questions about what they learned, to talk about the best part of the class, their assessment of the trainers and suggestions. We conducted a qualitative analysis to identify common themes and word clouds were created for two of the questions to display the frequency of responses. In addition, the top 12 most frequent responses of each question theme were explored in combination with participant demographics and offense types. The question: *"What, if anything, did you learn from this class?"* was chosen for this deeper analysis due to its ability to display what participants are learning in the program and thus potentially revealing areas for improvement. Findings include:

- Black participants were disproportionately *more* likely to mention learning to "walk away from certain situations";
- Hispanic participants were disproportionately *more* likely to mention learning to "not commit crime" and that "fighting is not the answer";
- Female participants were disproportionately *more* likely to mention learning that "fighting is not the answer" and to "listen to and understand others' perspectives"; and
- Participants who committed an offense against a person were disproportionately *more* likely to mention learning to "walk away from certain situations".

The final analysis conducted examined recidivism rates overall and by year, and then explored predictors of arrest using logistic regression. There are limitations to the arrest data (including a change in the data collection method and a lack of arrest dates which necessitated approximating dates of arrest from other variables available in the dataset, and re-arrest offense type) that require caution in overstating these results. Nonetheless, 82% of the 4,038 youth were not rearrested following participation in the workshop while 727 of 4,038 (or 18%) were rearrested. By year, the highest number of youth rearrested were from the 2013 cohort – 152 (or 31%) were arrested. Re-arrest within 1, 2, 3 and more than 3-years post-program indicates that among the 727 rearrested:

- 274 (38%) were rearrested within the first year after participation in the program;
- 218 (30%) were rearrested within the second year;
- 116 (16%) were rearrested within the third year; and
- 119 (16%) were rearrested three or more years post program participation.

Logistic regression which is used for the binary outcomes of arrest/no arrest, allows one to predict the outcome, while accounting for information contained in other variables which could explain that outcome. For example, male offenders are more likely to recidivate, thus one would want to "control" for gender in the analytic model. The final model included demographic, offense type, and key program variables including if the referral to GSMP was from the court, the number of days since the youth participated in the program, and the pre-test scores from Construct 4 – Agency/Potential scale. The model generates odds ratios, which were converted into predicted probabilities for ease of interpretation.

The results indicate that gender, offense, and key program variables significantly predict arrest among the GSMP participants. Males are 7% more likely to be arrested than females, while those charged a drug offense are 8% more likely to be arrested and youth who participated in the auto theft module are 12% more likely to be arrested. However, those charged with a weapons crime are 6% *less* likely to be arrested. Those who are referred from the court have a higher predicted probability of arrest of 3% (compared to those referred from another source) and pre-test Construct 4 Agency/Potential average score indicated that those with *lower* scores on the pre-test C4 are 7% *more* likely to be arrested and those with higher C4 pre-test scores.

Overall, the GSMP program is successful at both shifting youth perspectives and has an overall recidivism rate of 18% (727 of 4,038 rearrested).

Future evaluations should consider including a process evaluation to ensure that the program components are implemented as intended, and should seek official criminal history data to enhance the recidivism analysis. Finally, this program would benefit from the identification of another group of similarly situated youth that had no exposure to the program to be used in comparison to GSMP youth to assess if recidivism rates differ.

Overview

Choice Research Associates (CRA) was contracted to examine the characteristics and recidivism rates of youth referred to the Good Shepherd Mediation Program (GSMP).

GSMP offers services to youth who participate in the Juvenile Offender Diversion Program (JODP). This program is available to first-time offenders between the ages of 11 and 17 who have been arrested for a misdemeanor (e.g., underage drinking, loitering, disorderly conduct, defiant trespassing, harassment, vandalism, criminal mischief, terroristic threats, possession of a weapon on school property, shoplifting, auto-related offense) or a simple felony (e.g., assault). Referrals to the program come from the Philadelphia Police Department, Office of the District Attorney's Youth Aid Panels, judges in the Family Court Delinquency unit, probation officers, and the Philadelphia Juvenile Justice Services Center (formerly known as the Youth Study Center).[1]

GSMP also offers services to youth who participate in the Youth Delinquency and Violence Prevention (YDVP) program. This program serves youth between the ages of 11 and 17 who have not been arrested but instead have been identified as at-risk of delinquency. Youth are identified based on risky behavior that does not rise to the level of delinquency (e.g., truancy, peer associations, verbal or physical aggression) and are referred to GSMP by the Department of Human Services, community-based organizations, faith-based institutions, schools, and parents.

Youth in both JODP and YDVP are required to attend a conflict resolution workshop. This workshop is based on restorative-justice principles and designed to reduce recidivism. These programs offer at-risk and delinquent youth the opportunity to learn productive ways to manage conflict, control their anger, and make better decisions.

In addition to attending the conflict resolution workshop, youth who have been arrested for shoplifting and related offenses must also participate in the effects of retail theft module which addresses the personal and social pressures that can trigger retail theft and the hardships that shoplifters face in their lives, retail theft law and its consequences, and behavior modification techniques.

Youth who committed an auto theft offense which includes the unauthorized use of a motor vehicle and other auto-related offenses (e.g., keying a car, riding an ATV in the street), are required to participate in the impact of auto theft module during the conflict resolution workshop which focuses on the impact of auto theft related offenses on the youth, the victim, and the community. Youth are referred by the Youth Aid Panels, probation officers and Family Court judges. The auto theft module offers participants opportunities to use critical reasoning to enhance their positive decision-making skills. Activities include small and large group discussions, viewing an informative video on the effects of auto-related crimes on their lives and others; and the chance to speak with a member of the Philadelphia Police Department concerning the impact and ramifications of their crimes. Similarly, youth who committed a retail theft are required to participate in retail theft module during the conflict resolution workshop.

[1] In order to participate in the program, youth must admit involvement in the delinquent act.

Between 2001 and 2017 a total of 6,094 youth were referred to GSMP. While the age of eligible youth is 11 to 17 years, data show that youth as young as 7 and as old as 26 participated in the program.[2] The average age of participants was 15.2 years. Seventy-five percent of participants were Black/African American and 8% were white/Caucasian while 17% identified as another race (e.g., Asian, Pacific Islander etc.). The majority of attendees were male (66%) and were referred by Youth Aid Panels (69%) (See Appendix A for descriptive statistics for all youth who participated in GSMP between 2001 and 2017). Due to the lack of information regarding the survey questions used between 2001 and 2007, this analysis will focus on the 4,023 youth referred to GSMP between 2008 and 2017.

Data Issues

As mentioned, one limitation to the data analysis was the lack of information regarding the survey questions used between 2001 and 2007. While demographics for the youth participating in GSMP during these years could be examined, results of the program could not be explored. Missing data issues extended to the 2008-2017 data as well. For example, approximately 250 individuals in the program between 2008 and 2017 had an unknown offense type due to missing data.

Methodology

GSMP provided 25 spreadsheets containing participant workshop registration information and pre-post- test survey data for the periods from 2001 through 2017. The data were combined and based on youth name, gender, and year of attendance, unique identification numbers were assigned to each participant. For the recidivism data, GSMP provided 40 spreadsheets for youth who participated in the workshop.[3] These spreadsheets were reviewed and combined, and using the name of the participant, identification numbers were assigned to these data. The participant and the recidivism data were then converted into a SPSS (a statistical analysis software) and then combined, linking the files by the assigned identification number, to prepare this report.

Sample Descriptive

Table 1 presents the descriptive statistics for the youth who participated in GSMP between the years 2008 and 2017. The age of participants ranged from 7 to 26 years with a mean of 15.2 years old. 66% of participants were male. A majority of the participants were Black/African American (75%) and 14% of youth identified as Hispanic. Among these participants, 69% were referred by a Youth Aid Panel (YAP) and 29% were court ordered.

[2] Program staff advised that the offense must occur before the participant's 18th birthday. However, some participants may be older due to a gap between the offense and program attendance (or in the case of the 2 individuals over 21 years old, these may be data entry errors as neither had a date of birth listed) (GSMP Staff, Personal Communication, July 6, 2018).

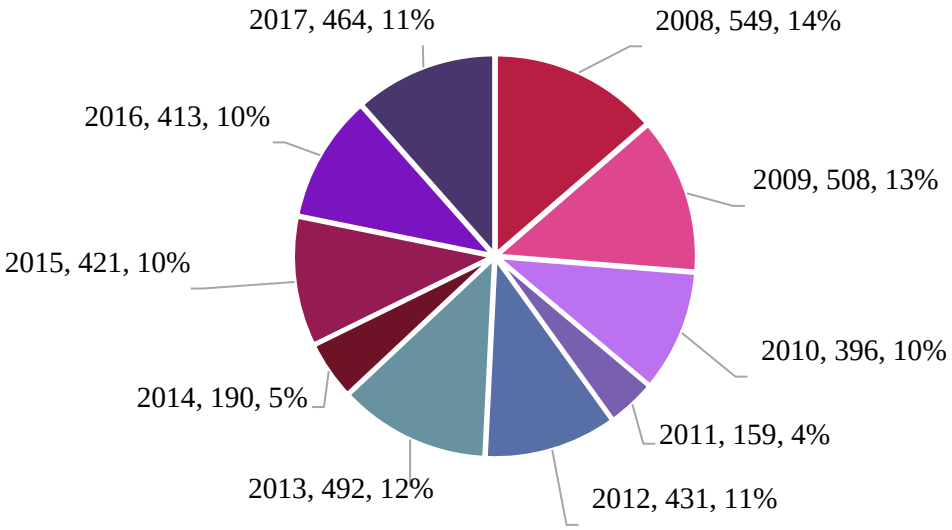
[3] Of those 40 spreadsheets, 16 were monthly spreadsheets from the District Attorney's office starting in 2015 through 2016. In addition, among these spreadsheets, some were participant rosters that did not have any recidivism data reported, others were draft or working copies of more complete spreadsheets.

Table 1: Demographics of Youth Referred to GSMP (2008-2017)

	N[4]	Freq.	Percent	Range	Mean (SD)[5]
Demographics					
Age	3,841			7 to 26	15.21 (1.81)
Race	3,837				
Black/African American		2,873	75%		
White/Caucasian		308	8%		
Other		656	17%		
Ethnicity	3,839				
Hispanic		552	14%		
Sex	3,891				
Male		2,553	66%		
Female		1,338	34%		
Referral Type	3,522				
Court Ordered		1,027	29%		
Community Referral		18	0.5%		
YAP		2,430	69%		
Other		47	1.3%		

Figure 1 provides the breakdown of GSMP attendees from 2008 to 2017. Attendance remained relatively consistent throughout the years with the exception of 2011 and 2014 where attendance dropped below 200 youth for each year.

Figure 1: Number and Percentage of Attendees by Year N=4,023



[4] N=Number of those with data available to assess.
[5] "Standard Deviation" indicates variation in the data. A larger SD more variation, smaller SD more consistency.

Table 2 provides the breakdown of offenses that youth referred to GSMP were charged with. While a small portion of youth were charged with multiple offenses, this analysis only utilized the first offense listed in the data. Individual offenses have been aggregated by type (Appendix B contains a breakdown of individual offenses). Person offenses include crimes such as assault, property offenses include crimes such as theft and vandalism; and public order offenses include crimes such as terroristic threats and disorderly conduct. Overall, 66% of youth referred were charged with a person (e.g., assault) or property (e.g., theft) offense. Offense type was missing for 279 youth.

Table 2: Offense Type of Youth Referred to GSMP (2008-2017) N=3,759

Offense Type	N[4]	Freq.	Percent
Aggregated Offense	3,759		
Person (assault, aggravated assault, robbery, harassment)		1,250	33%
Property (auto theft, theft, vandalism, burglary, trespassing)		1,226	33%
Drug (possession, use, and sales – all drug types)		459	12%
Weapon (possession of gun, knife, box cutter on school property)		521	14%
Public Order (terroristic threats, disorderly conduct, resisting arrest)		235	6%
Other (Conspiracy and Community Referrals (non-arrest)		68	2%

Taking a closer look at the 1,250 person offenses, Figure 2 provides a breakdown of the individual offenses. Simple assault (889 of 1,250 or 71%) and aggravated assault (254 or 20%) made up 91% of person offenses, while 82 youth charged with robbery accounted for 7% of person offenses. The remaining 2% of offenses ranged from harassment, intimidation and criminal recklessness to sexual assault and attempted murder.

Figure 2: Person Offense Type of Youth Referred to GSMP (2008-2017) N=1,250

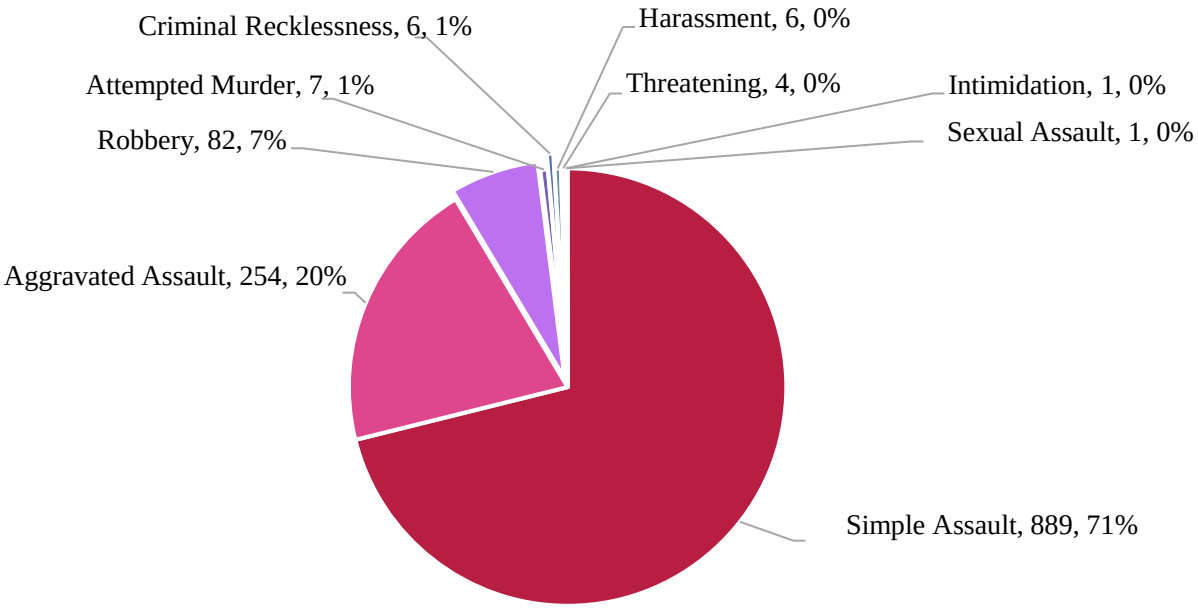


Table 3 presents the results from crosstabulations and one-way ANOVA analysis examining the relationship between demographics and the type of class youth participated – (e.g., auto theft and retail theft class modules or the conflict resolution class which involved youth charged with all offense types). Results indicate significant[6] differences across all demographic groups – although these differences may not be substantively informative. For example, while the average age for those in the auto theft module was 15.9 years, those in the retail theft class were on average 15.4 years old. Youth *not* associated with retail or auto theft offenses were the youngest with a mean age of 15.1 years ($p<.001$). The slight (albeit significant) difference between 15.9 vs. 15.4 vs. 15.1 years old is not particularly illuminating.

However, that is not the case with respect to the gender of participants in the three different workshops. Over 90% of youth in the auto theft module are male compared to 30% of those in the retail theft module ($p<.001$). Thus, there is a clear relationship between the type of offense and the gender of the participant. The conflict resolution workshop, which involves a range of other types of offenses, includes 68% male and 32% female participants.

Results also show the majority (over 70%) of youth across all three module-related offense types were African American/Black, followed by youth categorized as "Other" (including Latino, Asian, Indian, Caribbean, as well as those who identify as bi-racial and multi-cultural) and Caucasian/White ($p<.01$). Results also show significant variation across the three offense groups by Ethnicity, with Hispanics more likely to be associated with other offense types ($p<.05$). With respect to referral type, Table 3 shows significant variation with youth primarily coming from two sources -- Court Ordered and YAP. Among the retail theft group 88% were referred by YAP while 62% of the auto theft group were referred by YAP ($p<.001$).

Through discussions with GSMP staff, researchers became aware of a possible shift in the type of youth being referred to the program beginning in 2015. A school diversion program was started that diverted kids rather than arrest them. Consequently, the GSMP and Youth Aid Panel participation numbers dwindled, resulting in the referral of more serious offenders, like those who may have been involved in an aggravated assault. Staff was interested in seeing the portion of increase in more serious offenders since 2015, if this post-2015 group is different than the pre-2015 group as well as if the workshops have any positive impact on this group. In order to explore this issue, a variable was created ("Cohort") which differentiated 2015-2017 participants from 2008-2014 participants and was included in the construct and regression analysis. The results are discussed below.

[6]Differences that are statistically significant if the "p-level" indicator is $p<.05$ or below. This notation means that the findings are highly unlikely (e.g., for $p<.001$ - less than a 1 out of 100 chance or $p<.05$ less than 5 out of 100 chances) to be the result of chance or coincidence.

Table 3: Demographic Differences Across Type of Module/Workshop (2008-2017)

	Auto Theft Module		Retail Theft Module		Conflict Resolution (All Offense Types)	
	N[4]	Mean Age	N	Mean Age	N	Mean Age
Age***	342	15.9	418	15.4	2908	15.1
	N	%	N	%	N	%
Sex***						
Male	319	93%	126	30%	1995	68%
Female	26	7%	300	70%	951	32%
Race**						
African American/Black	241	70%	311	73%	2186	76%
Caucasian/White	40	12%	49	12%	212	7%
Other	63	18%	65	15%	498	17%
Ethnicity*						
Hispanic	57	17%	45	11%	430	15%
Referral Type***						
Court Ordered	125	38%	44	11%	786	30%
Community Referral	0	0%	2	<1%	14	<1%
YAP	205	62%	359	88%	1768	68%
Other	1	<1%	1	<1%	37	1%

Statistically significant levels *p <.05, **p<.01, ***p <.001

Table 4 presents the results from crosstabulations and one-way ANOVA analysis examining the relationship between demographics and offense types (person, property, drug, weapon, and Other). All demographic variables except Ethnicity (Hispanic) were found to significantly vary across the offense types.

For example, those with a weapon offense were on average the youngest with a mean age of 14.61 years where those with a drug offense were on average older with a mean age of 16.14 years (p<.001). Among youth with a drug offense, 86% were male and 14% were female. In comparison, 52% of those with weapon offenses were male and 48% were female (p<.001).

In regard to race, 79% of those with person or weapon offenses were African American/Black and while Caucasian/White youth were an overall small (8%) percentage of program participants, 14% of those with drug offenses are Caucasian/White. Finally, analysis also reveals significant variation across offense types by the referral source (p<.001). While 87% of youth associated with a drug offense were referred by YAP, 15% of youth were Court Ordered. In comparison among those with person related offense 52% were referred by YAP and 46% were Court Ordered. Overall the findings from Table 3 and Table 4 indicate significant demographic variations across the offense types (whether identified by module related or more general offense types).

Table 4. Demographic Differences Across Offense Type (2008-2017)

	Person		Property		Drug		Weapon		Public Order		Other	
	N[4]	Mean Age	N	Mean Age	N	Mean Age	N	Mean Age	N	Mean Age	N	Mean Age
Age***	1,214	14.99	1,194	15.36	452	16.14	518	14.61	231	15.20	59	15.31
Age Range	10 to 20		10 to 19		11 to 18		10 to 18		9 to 23		11 to 26	
	N	%	N	%	N	%	N	%	N	%	N	%
Sex***												
Male	767	62%	801	66%	389	86%	271	52%	166	71%	46	70%
Female	470	38%	411	34%	62	14%	246	48%	68	29%	20	30%
Race***												
African American/Black	958	79%	872	72%	290	66%	405	79%	169	73%	44	69%
Caucasian/White	61	5%	135	11%	62	14%	17	3%	18	8%	8	13%
Other	190	16%	201	17%	89	20%	92	18%	42	18%	12	19%
Ethnicity												
Hispanic	179	15%	161	13%	72	16%	77	15%	36	16%	7	11%
Referral Type***												
Court Ordered	491	46%	254	22%	47	12%	65	15%	75	34%	23	38%
Community Referral	7	<1%	4	<1%	1	<1%	1	<1%	0	0%	3	5%
YAP	559	52%	878	77%	347	87%	380	85%	141	64%	27	45%
Other	13	1%	9	<1%	4	1%	2	<1%	4	2%	7	12%

Statistically significant levels *p <.05, **p<.01, ***p <.001

Note: Age differences were analyzed using One-Way ANOVA and Post-Hoc Scheffe tests revealed significant age differences among many but not all of the subgroups. For example, those with drug offenses were found to be significantly older than those in the five remaining offense categories. On the other hand, those with Other offenses were only found to be significantly different (in this case younger) then those with drug offenses.

Overall Analysis and Results

Survey Completion and Attrition Analysis

Among 4,023 youth referred to GSMP, data were provided for 3,587 youth who completed a pre-test, a post-test, or *both* pre-and post-test. Among those 3,587, 97.99% (3,515) youth had pre-test survey data and 97.60% (3,501) had post-survey data.[7] An analysis of attrition was conducted to determine if the participants with pre-test data but did not have post-test data differed from the participants with both pre-test and post-test measures. If youth with post-test data differed from those without post-test data, it would suggest that the amount of change observed for participants with both pre- and post-test data may not be representative of the degree of change for all participants in the program. Appendix C presents the results from the attrition analysis which compares the pre-test scores of those who were post-tested to those who were not post-tested. No statistically significant differences were found between those who completed the post-test and those who did not.

Pre-Post Test Results

Youth who participated in GSMP completed a pre- and post-survey (See Appendix D and Appendix E). Both surveys provided youth with 23 statements to which they could select from a 5-point Likert scale assessing their level of agreement (1 = strongly disagree to 5 = strongly agree). Statements assess attitudes on physical violence, anger control, conflict resolution, understanding consequences of actions, helpfulness of the workshop, and others. Overall, 85% (3,429) of participants had data for *both* the pre- and post-survey.

In order to examine how GSMP may have impacted youth perceptions, a paired samples t-test was conducted. This type of analysis examined whether the mean difference between two sets of observations, in this case survey responses, are significantly different. Table 5 presents the results and shows significant changes in mean responses for 18 of the 23 survey questions.

It should be noted that for some of the statements (e.g., *It is ok to walk away from a fight whether or not you think you would win*) an increase in the average score would show a positive effect of the workshop, while a decrease in the average score for other statements (e.g., *Fighting usually solves a problem*) would show a positive effect of the workshop.

Measures with a statistically significant increase (meaning participants were (on average) ***more*** likely to agree with the statement post-workshop), included:

- It is okay to walk away from a fight whether or not you think you would win.
- Fighting usually solves a problem.
- I can control how I am feeling at any given time.
- I know what things people do or say that trigger my anger.
- When people make me angry I feel I am justified in acting out against them.

[7] According to program staff, a small portion of youth missing post-survey data were expelled from the class while most missing data can be attributed to lost forms. It is not likely that a youth failed to complete a post-survey since certificates are only given out once the test is turned in (GSMP Staff, Personal Communication, July 6, 2018).

- It's possible for me to think about the consequences of my behavior before I act on my feelings.
- I can choose different ways of reacting to someone when they make me mad.
- It's possible for people to see the same situation in different ways.
- It's important for me to understand where the other person is coming from before I do something in a conflict.
- What I want/need are more important than what the other wants/needs if there is a problem.
- I feel I should be in this program.
- I understand why I am in this program/I understand why I was in this program. (post)
- I think this workshop will be helpful to me/This workshop helped me. (post)
- I believe I will learn something that will benefit me/I learned something that will benefit me. (post)
- I see how my actions have affected my relationships with family, friends, and others.

Measures that saw a statistically significant decrease, (meaning participants were (on average) ***less*** likely to agree with the statement), included:

- Listening to the person you are angry with does not do any good.
- I understand the consequences that may happen to me if I commit another crime.
- I understand how my actions may affect others (family, friends, others) if I commit another crime.

Overall these results indicate that participants gained knowledge on avoiding physical violence, anger control, conflict avoidance, thinking of consequences, understanding how actions affect others, and an understanding of and appreciation for participation in the program.

However, some measures indicate some unexpected results. For example, it is surprising that participants were more likely to agree with the statement "*I see how my actions have affected my relationships with family, friends, and others*", but less likely to agree with the statement *I understand the consequences that may happen if I commit another crime"* after the program. This result may indicate that the program helped participants understand how their actions which led to their participation in the program affected loved ones, but not necessarily how a future criminal act may impact them directly.

Table 5: Pre-Post Test Outcomes N=3,429

Responses on a Scale of 1 to 5, where 1=strongly disagree to 5=strongly agree	N[4]	Pre-Test Mean	Post-Test Mean	Diff
Q1. It is okay to walk away from a fight whether or not you think you would win.	3,399	3.80	3.93	0.13***
Q2. Fighting usually solves a problem.	3,399	2.17	2.25	0.08***
Q3. I can control how I am feeling at any given time.	3,358	3.57	3.66	0.09***
Q4. I know what things people do or say that trigger my anger.	3,392	3.71	3.96	0.25***
Q5. When I'm angry I feel I must act on my anger right away.	3,395	2.50	2.54	0.04
Q6. Listening to the person you are angry with does not do any good.	3,370	3.22	3.00	-0.22***
Q7. When people make me angry I feel I am justified in acting out against them.	3,376	2.50	2.58	0.08***
Q8. It's possible for me to think about the consequences of my behavior before I act on my feelings.	3,398	3.71	3.93	0.22***
Q9. I can choose different ways of reacting to someone when they make me mad.	3,351	3.85	3.99	0.14***
Q10. It's possible for people to see the same situation in different ways.	3,325	3.85	3.93	0.08***
Q11. It's important for me to understand where the other person is coming from before I do something in a conflict.	3,328	3.73	3.83	0.10***
Q12. What I want/need are more important than what the other wants/needs if there is a problem.	3,144	2.81	2.93	0.12***
Q13. I feel I should be in this program.	3,228	2.77	3.44	0.67***
Q14. I understand why I am in this program/ I understand why I was in this program. (post)	3,231	3.95	4.18	0.23***
Q15. I think this workshop will be helpful to me/ This workshop helped me. (post)	3,208	3.63	4.16	0.53***
Q16. I believe I will learn something that will benefit me/I learned something that will benefit me. (post)	3,239	3.85	4.20	0.35***
Q17. I am responsible for the actions that brought me into this program.	3,248	4.16	4.17	0.01
Q18. I understand coming to this workshop is one of the consequences of my actions.	3,254	4.21	4.20	-0.01
Q19. I see how my actions have affected my relationships with family, friends, and others.	3,250	3.78	3.93	0.15***
Q20. I am willing to take responsibility for my actions that led to my arrest.	3,225	4.19	4.19	0.00

Responses on a Scale of 1 to 5, where 1=strongly disagree to 5=strongly agree	N[4]	Pre-Test Mean	Post-Test Mean	Diff
Q21. I realize how serious my crime could have been.	3,234	4.11	4.14	0.03
Q22. I understand the consequences that may happen to me if I commit another crime.	3,240	4.38	4.31	-0.07***
Q23. I understand how my actions may affect others (family, friends, others) if I commit another crime.	3,241	4.32	4.21	-0.11***

Statistically significant levels *p <.05, **p<.01, ***p <.001

Developing and Analyzing Survey Data with Latent Constructs

In the discussion above related to the survey results, each survey question was examined individually. This section explores the relationship among the 23 survey questions in order to examine the quality and usefulness of individual survey items as well as to determine the existence of underlying factors that can be used to create scales (otherwise referred to as latent constructs). In this process, we enhance both our data analysis and understanding of the results.

One of the benefits to using survey instruments to collect data is that it allows researchers to examine concepts that cannot be directly measured or adequately captured with just one question or item. In quantitative research these concepts are referred to as *latent constructs*. For example, a concept such as socioeconomic status may be best captured by multiple measures that can be combined to create a latent construct or scale. A socioeconomic status scale may include measures of income, level of education achieved, home value and/or other assets such as savings and investments.

In the case of the JODP program, GSMP staff identified 7 themes that potentially represent latent constructs: 1) Value in fighting, 2) anger management, 3) consequences of actions, 4) responsibility/accountability, 5) perspective taking, 6) competency, and 7) predicting future consequences. To test the existence of these themes, correlation, factor and reliability analysis were conducted.

The first analytic step, correlation analysis, shows the strength of the relationship between the 23 pre-survey test measures. Measures that correlate strongly with one another are good candidates for factor analysis. Results from the correlation analysis[8] show that all 23 pre-survey measures significantly correlate with one another with varying degrees of strength and direction (positive or negative relationships). There were also no indications of multicollinearity.[9] With confirmation that the measures significantly correlate with one another, a factor analysis was conducted.

[8] Results not presented but available upon request.

[9] Pearson's r values varied across the association tests between different measures however, the value never rose above .9 which is an indicator of multicollinearity. Multicollinearity occurs when two measures are almost perfectly correlated, hence a Pearson's r value above .9, and can be interpreted as these two items are essentially measuring the same thing. If not addressed (e.g., by removing one of the measures) the inclusion of these measures in an analysis can impact the results and lead to incorrect conclusions.

A factor analysis examines the response patterns across the 23 measures to identify how well the measures 'hang' together. Based on the response patterns associated with the measures, multiple factors can be identified and each measure (or item) with have a factor loading score. These factors loading scores range from -1 to 1, with 0 indicating no effect. The higher the value, the stronger the association.

If there is more than one factor derived from the data, each item will have a value associated with it, and a higher value on one factor indicates a better fit in comparison to the other factors. A general rule is that a value of .4 or above is a good indicator of adequate association with a factor. For example, if a factor analysis reveals two factors and an item scores .20 on the first and .50 on the second, there is a stronger association (better fit) with the second factor.[10]

Table 6 summarizes the results from the factor analysis[11] showing where the individual measures, or items, fit best across the four factors based on the "factor loadings" (see Appendix F). Seven items (Q17, Q18, Q19, Q20, Q21, Q22, Q23) loaded best on Factor 1, 5 items (Q2, Q5, Q6, Q7, Q12)[12] loaded best on Factor 2, 4 items (Q13, Q14, Q15, Q16) loaded best on Factor 3, and 7 items (Q1, Q3, Q4, Q8, Q9, Q10, Q11) loaded best on Factor 4. To further explore the strength of these measures as factors, reliability analysis was conducted. This type of analysis measures the internal consistency and dependability of these measures - before they are scaled together to create a construct.

Table 6 also shows the Cronbach's alpha level from the reliability analyses conducted on the grouping of measures across the four factors. Cronbach's alpha represents the level of internal consistency. While the threshold for weak, moderate, and strong internal consistency can vary across various researchers (and across disciplines), but for the current project we define Cronbach's alpha values above .7 as strong, values above .5 moderate and values .4 and below as weak. Based on these standards, the internal consistency of the four factors can be characterized as moderately strong to strong – findings that supported the creation of four latent constructs.[13]

---

[10] Note that as indicated in Appendix F, six measures (Q14, Q15, Q16, Q18, Q19, and Q21) had a factor loading score of .4 or greater on two different factors. While one value was always higher than the other, loadings above .4 on two factors may indicate a level of conceptual ambiguity, raising the question as to whether the measures are worded in a way that taps into two separate factors. Drawing on process of identifying conceptual labels discussed below, the development of an informative and useful survey not only depends on what the data is showing but also how the measures fit together conceptually, which is often a judgment call by practitioners and/or stakeholders. Thus, even when the data analysis indicates that a handful of measures "hang" together, if a common theme cannot be identified in the intent or wording of the question, then results from the data analysis are not informative.

[11] The Factor Analysis method used was Principle Axis Factoring with a Direct Oblimin Rotation.

[12] To simplify the findings of the scale analysis, these 5 variables were "reverse coded" to change the direction of the responses so that higher values would indicate more prosocial attitudes or more positive response in line with the other 3 scales or constructs examined.

[13] See Appendix G for the results of a paired samples t-test analysis examining the differences in mean scores between the pre-post surveys across the four latent constructs. Results show a statistically significant and positive increase in mean scores between the pre- and post-survey for Construct 2-Anger/Perspective ($p < .05$), Construct 3- Program Expectations ($p < .001$) and Construct 4-Agency/Potential ($p < .001$). In reflecting on the results from the paired samples t-test analysis of the 23 individual survey items presented in Table 5, these findings are not surprising. For example, four (Q17, Q18, Q20, and Q21) of the seven items used to create Construct 1-Actions/Consequences did not significantly change between the pre-post survey. Also, while the mean scores for Construct 2-Anger/Perspective were statistically significant between the pre- to post-period, the

The latent constructs, or scales, were then created by taking the average score from the respective items in each of the factors. [14]

The measures associated with each of the four factors were examined to identify common themes to create conceptual labels to represent the underlying constructs. In reviewing the questions associated with Factor 1, we identified a common thread in the reference to the youth actions, responsibility, and consequences. Based on this, Factor 1 was labeled "Construct 1- Actions/Consequences." However, Q21 (*I realize how serious my crime could have been)* stands out for two reasons. First, given the range of seriousness of offenses these youths committed to be referred to the class, they may have had equally variant recognition of the seriousness of the crime. Second, the wording seems somewhat different than the other six questions in this construct. Specifically, the other questions appear more self- or action- oriented -- using words such as "*I am responsible …"* or *"I see how my actions …*" or "*I am willing…".* This is an example where while the reliability and factor analysis results indicate that this measure fits well within Construct 1, it may be beneficial to explore revisions to the wording of this measure (or delete the question altogether) to increase conceptual cohesiveness.

In reviewing the measures that loaded best on Factor 2, three fit cohesively as all refer to feelings of anger. However, despite factor and reliability analysis indicating all five of these measures fit well together, there is a lack of obvious cohesiveness. Specifically, the fourth (Q2 *Fighting usually solves a problem*) and fifth measures (Q12 *What I want/need are more important than what the other want/needs if there is a problem*) do not appear to fit as well. Nonetheless, in order to represent all five measures, Factor 2 was labeled "Construct 2 - Anger/ Perspective." As with Factor 1, it may be beneficial to explore possible revisions to the wording of these measures.

The measures associated with the Factor 3 all clearly relate to the program or workshop (e.g., Q14 *I understand why I am in this program*). As a result, Factor 3 was labeled "Construct 3 - Program Expectations". Depending on how the individual survey questions are utilized by GSMP in their day to day functioning, it may be useful to explore the possibility of reviewing the 4 questions in Construct 3 to determine if one or more questions can be eliminated. If 2 or 3 questions can provide the same degree of understanding as 4 questions, then it is generally better to use fewer questions.

Finally, common threads found in the measures associated with Factor 4 were individual agency, potential to make prosocial decisions, and empathy -- the ability of the youth to put themselves in someone else's shoes. Based on this, this factor was labeled "Construct 4 - Agency/Potential." However, as with Factor 1 and 2, not all measures fit well together conceptually. For example, while six of the measures are structured from the perspective of the person completing the survey ("*I can…*" or "*It's possible for me to…*") Q10 refers to the ability of **people** to see the same

difference is small (0.03), particularly when compared to the difference in overall mean scores for Construct 3 and 4 (.44 and .14, respectively). This is possibly explained by the fact that Q6 *Listening when I'm angry I feel I must act on my anger right away* is the only question in Construct 2 that showed a substantial and significant reduction. The other 4 measures in this construct changed less and in the opposite direction – thus likely diminishing the impact of Q6.

[14] Factor loading results from pre-survey data were used to create the latent constructs for the post-survey data.

situation in different ways. While the factor and reliability analysis indicate that Q10 *is* a good fit with Factor 4, it may be beneficial to explore ways in which this measure can be revised to enhance its conceptual cohesiveness with the other measures associated with Construct 4.

The process of identifying common themes which tie measures together under one factor highlights one of the limitations of factor and reliability analysis. While this analysis can indicate which measures hang well together (conceptually consistent) and the strength of the construct, the analysis cannot necessarily tell you why these measures break out into these factors. There is both an art and science to developing strong and cohesive measures -- is not always easy and may require an iterative process. GSMP may wish to consider reviewing the survey questions, make revisions, collect additional data, and reconduct this analysis again to assess if the revisions strengthened the measure as a dimension of a larger construct. Another way to consider this issue is that while these results may not clearly indicate *how* measures should be revised, they provide a starting point – and with revision and reassessment, stronger measures can be developed.

Table 6: Results from Factor and Reliability Analysis of Pre-Test Survey Items

Factor	Construct Label	Items Combined (Average Score)	Cronbach's Alpha Level	Range	Mean	N
Factor 1	Actions/ Consequences	Q17, Q18, Q19, Q20, Q21, Q22, Q23	0.86	1 to 5	4.16	3,419
	Q17. I am responsible for the actions that brought me into this program.					
	Q18. I understand coming to this workshop is one of the consequences of my actions.					
	Q19. I see how my actions have affected my relationships with family, friends and others.					
	Q20. I am willing to take responsibility for my actions that led to my arrest.					
	Q21. I realize how serious my crime could have been.					
	Q22. I understand the consequences that may happen to me if I commit another crime.					
	Q23. I understand how my actions may affect others (family, friends, others) if I commit another crime.					
Factor 2	Anger/Perspective	Q2R, Q5R, Q6R, Q7R, Q12R	0.61	1 to 5	3.37	3,515
	Q2R. Fighting usually solves a problem.					
	Q5R. When I'm angry I feel I must act on my anger right away.					
	Q6R. Listening to the person you are angry with does not do any good.					
	Q7R. When people make me angry I feel I am justified in acting out against them.					
	Q12R. What I want/need are more important than what the other wants/needs if there is a problem.					
Factor 3	Program Expectations	Q13, Q14, Q15, Q16	0.79	1 to 5	3.55	3,422
	Q13. I feel that I should be in this program.					
	Q14. I understand why I am in this program.					
	Q15. I think this workshop will be helpful to me.					
	Q16. I believe that I will learn something from the program that will benefit me.					

Factor	Construct Label	Items Combined (Average Score)	Cronbach's Alpha Level	Range	Mean	N
Factor 4	Agency/ Potential	Q1, Q3, Q4, Q8, Q9, Q10, Q11	0.68	1 to 5	3.74	3,511
	Q1. It is okay to walk away from a fight whether or not you think you would win.					
	Q3. I can control how I am feeling at any given time.					
	Q4. I know what things people do or say that trigger my anger.					
	Q8. Its possible for me to think about the consequences of my behavior before I act on my feelings.					
	Q9. I can choose different ways of reacting to someone when they make me mad.					
	Q10. It's possible for people to see the same situation in different ways.					
	Q11. It's important for me to understand where the other person is coming from before I do something in a conflict.					

R=Reverse coded so that the all the items in the scale are in the same direction (e.g., higher values = more prosocial attitudes).

The next section explores demographic differences across the four Factors.

Demographic Differences by Construct

Not only does the use of factor analysis allow for the examination of concepts that are best measured by more than one or two items, but it also reduces data to simplify the analyses. For example, instead of examining demographic differences across 23 individual pre-survey items, differences can now be analyzed using the four constructs -- Construct 1 Actions/ Consequences (C1); Construct 2 Anger/Perspective (C2); Construct 3 Program Expectations (C3); and Construct 4 Agency/Potential (C4). In addition, if statistically significant differences are found, then this informs the next step in the analysis – multivariate regression - where we would include these variables in the model to ensure that these factors are accounted for when explaining the outcomes.

This section presents the results from one-way ANOVA analyses and independent samples t-tests that were used to examine the differences in the average pre-survey scores by construct across the following demographic measures.[15]

[15] We use several types of statistical methods and tests to determine if differences are statistically significant. We use one-way ANOVA (such as reported in Table 7) to determine if the means on particularly measures of interest between 3 or more groups are equal (e.g., are the mean scores pre-and post-test different for youth based on the referral type – court ordered, community referral, YAP, or Other.) We use the F-statistic (indicated in the text and the parenthesis in Table 7 below) to compare means of 3 or more groups, and the independent samples t-Test to compare the means of 2 independent groups (such as the results provided in Table 8**Error! Reference source not found.**). Both test statistics the F and t-Test, are designed to determine if the mean scores are *significantly* different.[6] Finally, if there is a significant difference in the examination of 3 or more groups (as indicated by the F-Test), we conducted the post hoc Scheffe test to determine *which* pairs of means were significantly different from one another. For example, with a 3-category race variable involving African American, Caucasian, and Other, the mean score on a measure for African Americans would be compared to Caucasians and Other. The mean score for Caucasians would be compared to African Americans and Other. The mean score for Other would be compared to the mean scores for African Americans and Caucasians.

The measures were categorized as follows:

- Race: African American/Black, Caucasian/White and Other;
- Offense Related Module/Workshop: retail theft, auto theft, and other offense;
- Referral Source: Court Ordered, Community Referral, YAP, and Other; and
- Offense Type: Person, Public Order, Property, Drug, Weapon, and Other.

Race

As indicated in Table 7, the overall ANOVA results examining the mean score differences across 3 race groups was statistically significant for 3 of the 4 constructs (C1, C2, and C4):

- Construct 1-Actions/Consequences (F=11.46 (2,3311) p <.001)[16];
- Construct 2-Anger/Perspective (F=9.25 (2,3389) p <.001); and
- Construct 4-Agency/Potential (F=5.28 (2,3385) p <.01).

Results indicate that while there was no significant difference between African Americans and youth identified in the "Other" category, Caucasians reported significantly higher pre-test mean scores than African Americans or those in the Other race category on these 3 constructs.[17] For example, for C1 – Actions/Consequences, while youth of differing races all scored in the "agree" range in the pre-test survey (as evidenced by an overall average score of 4.16 on the items contained in this construct), the 273 Caucasians had a significantly higher average score of 4.35, compared to the 2,478 African Americans with an average score of 4.15 or those 563 who define as "Other" race with a score of 4.12.

Similarly, for C2 – Anger/Perspective, the overall average pre-test score is 3.37 – within the "neither agree nor disagree" range. But for 275 Caucasians, their average score was 3.54 – significantly different from African Americans or those of "Other" races (an average of 3.35 and 3.37, respectively).

Module/Workshop

Overall ANOVA results examining Offense Related Modules vs. Workshop offenses were statistically significant for all 4 constructs, as follows:

- Construct 1-Actions/Consequences (F=30.48 (2, 3198) p <.001);
- Construct 2-Anger/Perspective (F=10.78 (2, 3271) p <.001);
- Construct 3-Program Expectations (F=25.58 (2, 3200) p <.001); and
- Construct 4-Agency/Potential (F=8.12 (2,3267) p <.001).

[16](F=11.078 (2,2311) p <.001) breaks down as follows: F= is the F-statistic, (2, indicates the degrees of freedom (DF) in the model, and 2311) is the N or number of cases with data included in the analysis. The p<.001 is the p-value or level of statistical significance for this model.

[17]Significance between pairs is identified with a superscript letter of the associated group. For example, A for African American, C for Caucasian, and O for Other. To clarify where groups start with the same letter, additional letters were added for clarification. For example, Person offense type is indicated by Pe, Public Order by Pu, and Property by Pr.

For Construct 1-Actions/Consequences and Construct 3- Program Expectations, those with retail theft offenses reported significantly higher mean of 4.45 than both the auto theft (4.19) and other offense (4.15) groups. Constructs 2-Anger/Perspective and 4-Agency/Potential had similar patterns -- attendees in the other offense group reported significantly lower mean scores (average score of C2 - 3.34 and C4 - 3.73) than those in those with an auto offense (C2 - 3.47and C4 - 3.86) or retail theft offense (C2 - 3.49 and C4 - 3.83).

For Construct 3-Program Expectations, which had an overall average of 3.55 (falling on the scale between neither agree nor disagree and agree) revealed that retail theft had a significantly higher average of 3.86 compared to those charged with auto theft (3.63) or other offenses (3.52).

Referral Source

Overall ANOVA results examining the constructs by type of referral were statistically significant for 3 of the 4 constructs, as follows:

- Construct 1-Actions/Consequences ($F=21.18$ (3,2957) $p < .001$);
- Construct 3-Program Expectations ($F=31.18$ (3,2959); and
- Construct 4-Agency/Potential ($F=3.55$ (3,3041) $p < .05$).

For both Construct 1-Actions/Consequences and Construct 3-Program Expectations, youth referred by YAP to GSMP had significantly higher average scores in this area prior to beginning the workshop than either Court Ordered or Other Referral Types. Finally, while the overall model for Construct 4-Agency/Potential was significant with $p < .05$, there was no reported Scheffe differences between referral groups.[18]

Offense Types

The ANOVA results reveal that 5 of the 6 offense types were statistically significant across one or more constructs.

- Construct 1-Actions/Consequences ($F=14.10$ (5,3195) $p < .001$);
- Construct 2-Anger/Perspective ($F=7.54$ (5, 3268) $p < .001$);
- Construct 3-Program Expectations ($F=11.78$ (5, 3197) $p < .001$); and
- Construct 4-Agency/Potential ($F=9.35$ (5, 3264) $p < .001$).

For Construct 1-Actions/Consequences, youth referred to the program for a person related offense reported significantly lower mean scores than those with property, drug or weapon offenses. This pattern is also evident in Construct 4-Agency/Potential. The overall average score for Construct 4 for all youth irrespective of offense type was 3.74 – falling between "neither agree or disagree" and "agree". When viewed by offense type, youth with person offenses start with significantly lower scores on this construct than those charged with property, drug or weapons crimes (3.66 vs. 3.81 and above for the other offenses). Those charged with

[18] We also ran a TUKEY post hoc test to account for the unequal sample sizes across the referral type. We found that YAP referrals were significantly different from the "other" referral source at a p-value of $p<.08$. We are only reporting significant findings of $p<.05$.

public order crimes (including disorderly conduct and resisting arrest) also report significantly lower scores (average of 4.05 on C1) than those charged with property crimes (average of 4.30).

For Construct 2-Anger/Perspective, those with person related offenses reported significantly lower mean scores than property offenders (3.28 vs. 3.45) and Construct 3-Program Expectations, property related offenses reported a significant higher average pre-test score of 3.72 than those person and drug offenders (3.46 and 3.49, respectively).

Table 8 presents the results from independent sample t-tests examining the demographic differences in average pre-test scores across the four constructs. These measures were coded as dichotomous variables representing group membership[19] and include:

- Ethnicity (Hispanic, non-Hispanic);
- Gender (Male, Female); and
- Cohort (cohort 1: 2008-2014, cohort 2: 2015-2017).

Ethnicity

T-tests examining ethnicity revealed a significant difference in the average pretest survey in Construct 4-Agency/Potential ($t=2.17$ (3390) $p<.05$) [20]. While youth overall are in the "agree" range on C4, Hispanic youth reported lower mean scores than non-Hispanics (4.12 vs. 4.18).

Gender

We also observe significant gender differences with 2 of the 4 constructs:

- Construct 1-Action/Consequences ($t=-2.57$ (3330) $p<.05$); and
- Construct 3-Program Expectations ($t=-2.25$ (3333) $p<.05$).

On C1 and C3, female youth report higher mean scores than males. Specifically, for C1, females begin the workshop with an average of 4.21 versus males of 4.14. Similarly, female students score higher program expectations than male students – 3.60 vs. 3.53.

Cohort

Finally, there were differences based on the cohort year youth participated in the workshops. T-test analysis revealed significant differences in 2 constructs:

- Construct 2-Anger/Perspective ($t=4.08$ (3528) $p<.001$); and
- Construct 4-Agency/Potential ($t=2.26$ (3524) $p<.05$).

[19] If the youth identifies as Hispanic, the ethnicity variable is coded as 1 for "Yes" and if not Hispanic, it is coded as 0 for "No". Likewise, male youth are coded as 1 and female as 0; and those in the later cohort (2015-2017) are coded 1 and the earlier cohort is 0.

[20] ($t=2.17$ (3990) $p<.05$) breaks down as follows: t= is the t-statistic and 3390) is the N or number of cases with data included in the analysis. The $p<.05$ is the p-value or level of statistical significance for this model.

Youth who participated in Cohort 2 (from 2015 to 2017) had lower average pre-test scores on than Cohort 1 (participants from 2008 to 2014) on both constructs. Cohort 2 youth scored 3.30 on C2 (vs. Cohort 1 3.40) and 3.71 on C4 (vs. Cohort 1 score of 3.76). These results may provide support for the program staff's concerns that youth participating after 2015 may be qualitatively different than those pre-2015. GSMP staff observed that youth in Cohort 2 were repeat offenders who may have been involved in more serious offenses (e.g., aggravated assault).

Taken together these results clearly indicate that different groups of youth - whether based on race, ethnicity, gender, referral or offense type - do not have the same "starting point" when they first come to the GSMP workshop. For example, female youth have higher pretest-survey scores with more positive scores than male participants. These differences should be considered when updating or revising the GSMP curriculum or workshop approach. It may be beneficial to create tailored content for different groups.

This information begs the question: If different groups of youth come into the workshop with significantly different perceptions (or starting points), do these factors also impact the level of change in their survey outcomes? To explore this issue, the next section of this report presents the findings from a regression analysis examining if demographics and offense characteristics predict the level of change in pre-post mean scores across the four constructs.

Table 7: Pre-Test Demographic Differences Across Four Constructs (ANOVA Results)

	C1: Actions/ Consequences		C2: Anger/Perspective		C3: Program Expectations		C4: Agency/ Potential	
By Race	F=11.46 (2,3311)[16] ***		F=9.25 (2,3389) ***		F=1.83 (2,3314)		F=5.28 (2,3385) **	
	N[4]	**Mean**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**
AA/Black[A]	2,478	4.16[C]	2,537	3.35[C]	2,481	3.55	2,534	3.74[C]
Caucasian/ White[C]	273	4.35[A,O]	275	3.54[A,O]	273	3.65	275	3.85[A,O]
Other[O]	563	4.12[C]	580	3.37[C]	563	3.55	579	3.70[C]
	C1		**C2**		**C3**		**C4**	
Module or Workshop	F=30.48 (2,3198) ***		F=10.78 (2,3271) ***		F=25.58 (2,3200) ***		F=8.12 (2,3267) ***	
	N	**Mean**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**
Auto Theft[Au]	316	4.19[R]	323	3.47[Ot]	316	3.63[R]	322	3.86[Ot]
Retail Theft [R]	354	4.45[Au,Ot]	360	3.49[Ot]	354	3.86[Au,Ot]	360	3.83[Ot]
Other Offenses[Ot]	2,531	4.15[R]	2,591	3.34[Au,R]	2,533	3.52[R]	2,588	3.73[Au,R]
	C1		**C2**		**C3**		**C4**	
Referral Type	F=21.18 (3,2957) ***		F=0.36 (3,3045)		F=31.18 (3,2959) ***		F=3.55 (3,3041) *	
	N	**Mean**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**
Court Ordered [Cou]	878	4.04[Y]	908	3.35	879	3.35[Y]	907	3.72
Community Ref	15	4.08	17	3.36	15	3.43	17	3.58
YAP[Y]	2,026	4.24[Cou,Oth]	2,081	3.38	2,027	3.67[Cou,Oth]	2,078	3.77
Other[Oth]	42	3.90[Y]	43	3.35	42	3.28[Y]	43	3.54
	C1		**C2**		**C3**		**C4**	
Offense Type	F=14.10 (5,3195) ***		F=7.54 (5,3268) ***		F=11.78 (5,3197) ***		F=9.35 (5,3264) ***	
	N	**Mean**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**
Person[Pe]	1,056	4.07[Pr,D,W]	1,089	3.28[Pr]	1,057	3.46[Pr]	1,088	3.66[Pr,D,W]
Public Order	179	4.05[Pr]	183	3.36	179	3.53	183	3.72
Property[Pr]	1,075	4.30[Pe,Pu]	1,095	3.45[Pe]	1,075	3.72[Pe, D]	1,093	3.81[Pe]
Drug[D]	392	4.24[Pe]	399	3.42	393	3.49[Pr]	399	3.81[Pe]
Weapon[W]	449	4.20[Pe]	456	3.38	449	3.57	455	3.84[Pe]
Other	50	4.07	52	3.22	50	3.38	52	3.70

F= F-Test statistic, (#, = Degrees of Freedom (DF) and ####) = N (Number of cases included in the test of the model)
Statistically significant levels *p <.05, **p<.01, ***p <.001

Table 8: Pre-Test Demographic Differences Across Four Constructs (t-Test Results)

	C1: Actions/Consequences		C2: Anger/Perspective		C3: Program Expectations		C4: Agency/Potential	
Ethnicity	t=1.631 (3316)		t=.34 (3394)		t=.18 (3319)		t=2.17 (3390)*	
	N[4]	**Mean**	**N**	**Mean**	**N[4]**	**Mean**	**N**	**Mean**
Non-Hispanic	2,837	4.18	2,901	3.37	2,840	3.56	2,898	3.76
Hispanic	479	4.12	493	3.36	479	3.55	492	3.69
	C1		C2		C3		C4	
Gender	t=-2.57 (3330)*		t=1.43 (3408)		t=-2.25 (3333)*		t=-.42 (3404)	
	N	Mean	N	Mean	N	Mean	N	Mean
Male	2,184	4.14	2,241	3.38	2,186	3.53	2,239	3.74
Female	1146	4.21	1,167	3.34	1,147	3.60	1,165	3.75
	C1		C2		C3		C4	
Cohort	t=0.26 (3431)		t=4.08 (3528)***		t=1.03 (3434)		t=2.26 (3524)*	
	N	Mean	N	Mean	N	Mean	N	Mean
Cohort 1 (2008-2014)	2,175	4.16	2,246	3.40	2,176	3.56	2,242	3.76
Cohort 2 (2015-2017)	1,244	4.16	1,269	3.30	1,246	3.53	1,269	3.71

t= t-Test statistic and ( ) = N (Number of cases included in the analysis)
Statistically significant levels *p <.05, **p<.01, ***p <.001

Regression Analysis

Ordinary Least Squares (OLS) regression is the most appropriate method of analysis to explore the degree of change in the average pre-to post-test survey scores while accounting for information contained in other variables which could explain that outcome (e.g., older youth may be less likely to report the workshop experience as positive as younger youth), thus one would want to "control" for age in the analytic model). This method is advantageous due to the ability of the model to combine a number of different variables and assign a relative weight to each so that the effects of the independent variables on the dependent variable allows for the "unique contribution" of each.[21]

The regression results presented in Table 9 examine how demographic and offense characteristics impact the degree of change expressed by workshop attendees based on average pre- to post-scores across each of the constructs. This table includes four models, each one representing the four latent constructs.

All four models include the same demographic and offense measures. The prior analysis which found significant demographic differences in the constructs informed our selection of the variables to include in the regression model.[22] Specifically, as Caucasian youth had significant differences from African American or "Other", we included a variable called "White" into the model where if a youth was White, they were coded as 1, and those of any other race were coded as 0.

Besides age, a continuous variable, other variables were coded into dichotomous measures. Gender has 2 categories – so males were coded as 1, and females as 0. Categorical variables with 3 or more types (race, offense, and referral type) were recoded into a dichotomous (or "dummy") variables. For example, there are six categories of offenses and each one was recoded into its own dichotomous variable (1 = person, 0 = all other offenses). The final step in conducting a linear regression required designating one category as the reference group for the offense types. In this case, we include property offenses as the reference group.

The results presented in Table 9 show the variance explained by each of the four models is small (see the adjusted r-squared values) – with the variables in the model only explaining from less than 1% to 2% of the outcomes. This is not entirely surprising given that typically, studies of human behavior are complex – and there may be any number of factors that are not captured in our data that may influence the outcomes.[23] In addition, the number of cases included in the regression dropped from over 3,300 to less than 2,700 because the analysis automatically removes any case that is missing one or more variables in the model (e.g., over 500 cases were dropped because we are missing the referral type).[24] Nonetheless, the results of the regression

[21] Allison, P.D. (1999). Multiple Regression: A Primer Thousand Oaks, CA: Pine Forge Press. P. 3.

[22] Initially we also included the type of module/workshop (auto and retail theft vs all others) but ultimately omitted those variables because the overall offense type captured that information.

[23] Including, but not limited to differences in the program implementation over time.

[24] This method of adjusting for missing data is referred to as "list-wise". The advantage to this method is that by using only the cases that have complete data, we can explore the relative weight of each different variable because only the *same* youth are analyzed. We also explored these models using "pair-wise" where all cases were included even if they were missing one or more of the variables. The results were substantively similar to the

analysis are informative. By conducting this analysis, we are in a better position to explain what influences the degree of change reported by GSMP workshop youth on the 4 constructs. Table 9 reports both the unstandardized coefficient (*b*)[25], the standard error (*SE*) [26] and the standardized beta coefficient (β). We are primarily reporting on Beta (β) in this report because this statistic compares the *strength* of the impact of each individual independent variable to the dependent variable, with all the other variables held constant. The values represent a correlation which ranges from 0 to 1 (0% to 100%), and the higher the absolute value of the beta coefficient, the stronger the effect. While the statistically significant Betas across the models are low overall (ranging from .04 to a high of .09) this still provides a guide to the most influential factors.

Looking at each construct separately, we see 6 variables are significant predictors -- age of youth, being male, white, or a person, drug, and weapon offender all significantly impacts the degree of change from pre-to-post test on C1 – the Actions/Consequences scale. Older youth exhibit *more* change on this scale, as do those charged with a person, drug or weapon offense. Gender and race have a negative relationship with the scale – so being male and white youth have less change than female or non-white youth. Of these statistically significant factors, looking at the Beta (β), we note that relative to the other variables, those charged with a person offense has the strongest impact on C1 (at .08); followed by older youth, and those charged with a weapon offense (both .07).

For C2 – Anger/Perspective – there are 4 predictor variables that are statistically significant – age of youth, male, person and weapon offender with males showing the strongest impact. Older youth, those charged with a person or weapon offense score have a positive impact on C2, while if you are male, you are *less* likely to report an improvement in perceptions on the C2 scale.

Exploring C3 – Program Expectations – 6 variables are significant and all are positively related to the outcome – age, person, drug, weapon, and public order offenders, and those referred by the court. Here person and weapon offenders both have a β of .09 – indicating the strongest influence on reporting a change from pre- to post on this scale.

Finally, on C4 – Agency and Potential – there are only 2 significant predictor variables – Male (negative relationship with C4) and person offenders (positive relationship with C4). However, the overall model was not statistically significant.

Overall, across the 3 significant scales (C1, C2, C3), we note that the key factors are the age of youth, gender, and type of offense (with more serious offenders showing the greatest impact on these outcomes). Notably, when these factors were included in the model, cohort (those in the program from 2015-2017 compared to those from 2008 to 2014) was not a significant predictor of improvement in perceptions across these 4 constructs.

---

list-wise regression models, however, because the model now includes all cases that have data (thus each variable contains *different* cases) we are not able to discuss the degree to which each variable influences the model relative to other variables.

[25] The unstandardized coefficient (*b*) represents the slope of the line between the predictor variable and the variable of interest (e.g., so for every one-unit increase in a variable, the dependent variable increases or decreases by the amount. For example, for each year older, the difference in Construct 1 increases by .03.

[26] The standard error for *b* (*SE*) represents the degree of variation in the variable, similar to a standard deviation of the mean.

Table 9: OLS Regression Predicting Differences in Mean Scores Across Four Constructs

Beta Coefficient *(B)* Standard Error (*SE*) and Standardized Beta (β)	**Model 1*** C1: Actions/ Consequences**		**Model 2*** C2: Anger/ Perspective**		**Model 3*** C3: Program Expectations**		**Model 4[NS] C4: Agency/Potential**	
Independent variables	*b (SE)*	β	*b (SE)*	β	*b (SE)*	β	*b (SE)*	β
Age of Youth	.03 (.01)**	.07	.03 (.01)**	.07	.03 (.01)**	.06	.01 (.01)	.02
Male (Yes=1; No=0)	-.08 (.03)**	-.06	-.15 (.03)***	-.09	-.07 (.04)	-.04	-.05 (.03)*	-.04
Hispanic (Yes=1; 0=No)	-.05 (.04)	-.03	.01 (.04)	.00	-.07 (.05)	-.03	.00 (.04)	.00
White (Yes=1; 0=No)	-.11 (.05)*	-.05	.01 (.05)	.00	-.11 (.06)	-.03	-.04 (.04)	-.02
Person[a] (Yes=1; 0=No)	.12 (.03)***	.08	.09 (.04)*	.06	.17 (.04)***	.09	.09 (.03)**	.07
Drug[a] (Yes=1; 0=No)	.10 (.05)*	.05	.08 (.05)	.04	.20 (.06)**	.07	.02 (.04)	.01
Weapon[a] (Yes=1; 0=No)	.13 (.04)**	.07	.14 (.05)**	.06	.24 (.06)***	.09	.03 (.04)	.02
Public Order[a] (Yes=1; 0=No)	.08 (.06)	.03	-.03 (.06)	-.01	.17 (.08)*	.04	.06 (.05)	.02
Other Offense[a] (Yes=1; 0=No)	.18 (.12)	.03	.04 (.13)	.01	.20 (.16)	.03	.13 (.11)	.02
Court Referral (Yes=1; 0=No)	.03 (.03)	.02	-.01 (.03)	.00	.11 (.04)**	.06	.01 (.03)	.01
Cohort 2 (2015-2017=1)	-.02 (.03)	-.01	.05 (.03)	.03	-.04 (.04)	-.02	-.02 (.03)	-.02
Constant	-.41 (.12)**		-.41 (.13)**		-.05 (.16)		.07 (.11)	
Observations	2,554		2,668		2,556		2,666	
Adjusted R-squared	0.015		0.014		.018		.003	

Statistically significant levels *p <.05, **p<.01, ***p <.001 [NS] = Not Significant
VIF<2.5, indicating no multicollinearity
[a] Property Offense is reference category

Pulling together the findings from this section it is clear that while the 23 survey measures fit well together into four constructs, there is still room to strengthen or fine tune the measures in how and what construct they are intended to capture. In addition, while different types of youth (gender, offense type etc.) may come to the class with different "starting points," pre-post analysis reveal that in general (3 out of the 4 constructs) their perceptions improve between the pre- to post-survey. In turn, the regression analysis takes this a step further by controlling for significant factors such as age, gender, race, and offense, and reveals that participation in the workshop renders a significant change (in the desired direction) in the perspectives of youth in the areas of Actions/Consequences; Anger/Perspective; and Program Expectations. The lack of significance in the regression model for Agency/Potential suggests that either these measures or the program may need to be adjusted to produce the desired change. This is an area for further program development.

Qualitative Data

In addition to the Likert scale questions, participants had the opportunity to respond to six narrative questions in the post-survey. These questions include the following:

- *What, if anything, did you learn from this class?*;
- *What was the best part of the class?*;
- *What would you change about the class (aside from how long it lasted)?*;
- *What did you think about the trainers?*;
- *Do you have any suggestion for the trainers?*; and
- *Please add any comments that you would like to share here*.

The answers provided were reviewed to identify common themes and word clouds were created for two of the questions to display the frequency of responses.[27] In addition, the top 12 most frequent responses of each question theme were explored in combination with participant demographics.

In response to the question *What, if anything, did you learn from this class?* participants who responded most often mentioned "conflict resolution skills", "thinking before taking action", and "anger management skills" (se). Other common responses included learning to "walk away from certain situations", to "not commit crime", the "consequences of actions", "listening to and understanding others' perspectives", and that "fighting is not the answer" when it comes to resolving conflict. Overall, participants expressed learning many valuable lessons and alternative methods of handling situations that may arise in their future.

This question was also chosen for deeper analysis due to its ability to display what participants are learning in the program and thus potentially revealing areas for improvement. Top responses to this question were explored in combination with participant demographics and offense types. These findings may or may not be helpful in reviewing the curriculum and/or revisions to the pre/post survey, nonetheless, interesting findings include:

[27] The word clouds examine individual words as opposed to complete phrases. Additionally, common words such as "the" and "before" are excluded.

- Black participants were disproportionately *less* likely to mention learning "consequences" and to "not commit crime", but disproportionately *more* likely to mention learning to "walk away from certain situations";
- Hispanic participants were disproportionately *more* likely to mention learning to "not commit crime" and that "fighting is not the answer" when it comes to resolving conflict, but disproportionately *less* likely to mention learning to "walk away from certain situations";
- Female participants were disproportionately *less* likely to mention learning to "not commit crime", but disproportionately *more* likely to mention learning that "fighting is not the answer" and to "listen to and understand others' perspectives"; and
- Participants who committed an offense against a person were disproportionately *less* likely to mention learning "consequences" and to "not commit crime", but disproportionately *more* likely to mention learning to "walk away from certain situations".

Figure 3: "What Did You Learn?" Word Cloud

The most common response for the *What was the best part of the class?* question was the role playing/skit activity followed by the introduction part of the class where participants shared their stories and had the opportunity to learn about their classmates. The agree/disagree activity and the movie were also commonly mentioned as being favorite parts of the class. Overall, participants seemed to very much enjoy the activity portions of the workshop. In addition to the role play/skit and agree/disagree activities, participants also often mentioned enjoying the

activities in general and the human knot and triggers activities specifically. Participants also enjoyed the opportunity to meet new people and share their opinions with others.

Less than half of participants provided a response to *What would you change about the class (aside from how long it lasted)?* (see Figure 4). Participants most often mentioned that they would like for the class to be shorter despite the instruction to not comment on how long it lasted. The next most common responses indicated that participants would like more activities, more breaks, and different food for lunch. Some responses indicated that different activities and more fun in the class were desired. Interestingly, many respondents used this question to reflect on their own prior behavior, indicating that they would change how they acted before attending the class.

Figure 4: "What Would You Change?" Word Cloud

Responses to the question *What did you think about the trainers?* were overwhelmingly positive. The most common response was a simple indication that the participant liked the trainers. Other common responses indicated that participants thought the trainers were nice, great, helpful, and fun. Additionally, many respondents indicated that the trainers were respectful, smart, down to earth, and understanding. The very few negative responses most often simply indicated that the participant disliked the instructors or felt that they were boring. Overall, however, the trainers during this time period were well received.

Approximately twenty percent of participants provided a response for the question *Do you have any suggestion for the trainers?* Respondents who had a suggestion most often said for them to keep up the good work or indicated that they were good trainers and helpful. As far as common specific suggestions, respondents said to have more activities, be more open with the class and be stricter with students who act out during class. Many students expressed some frustration with interruptions to the workshop by misbehaving participants. Additionally, several respondents indicated that they would like to be able to stay connected with the trainers after the workshop.

Additional comments provided by respondents most often expressed thanks and indicated that they enjoyed the program. Furthermore, many respondents indicated that the program was helpful and that the instructors did a good job teaching the class. Additionally, many students expressed a desire to return as guest speakers for future classes. Finally, several participants used this question to indicate that they were going to change their behavior as a result of participating in the workshop.

Recidivism Analysis Overall Outcomes

As indicated in the methodology section, GSMP provided 40 spreadsheets with re-arrest data youth charged who attended the workshop from the period of 2008 to 2017. In total there were 4,038 attendees in the period from 2008 to 2017.

The types of information available and consistency of data varied across the 40 spreadsheets. For example, in some cases the data included an indicator of a juvenile arrest, and in some cases, an adult arrest, but there were few dates of arrest. Among the 727 individuals with a juvenile and/or adult arrest, originally 274 (38%) had a date of arrest, and 288 (or 40%) had offense information. This is relevant because without dates of arrest we cannot determine how long was the period from the workshop to the arrest. Consequently, we are unable to precisely account for the time "at risk" to recidivate. For example, those who participated in the workshop in 2013 had a longer period to engage in criminal behavior and to have that behavior come to the attention of the criminal justice system than those who participated in the workshop in 2017.

In addition, there was no indicator of when the data was collected. This is important because once youth reach the age of 18, GSMP no longer receives recidivism data. It is possible that those in the earlier years aged-out of the data collection effort prior to recidivating. Conversely, there has been long-standing evidence of a relationship between age and criminal behavior -- adolescents engage in antisocial behavior on an increasing basis from the age of 15 until they peak at age 19.[28] After 19, there is a sharp decline.

Finally, beginning around 2015, the recidivism data collection process changed. Prior to that time, GSMP staff were provided access to look up each youth and record recidivism activity including date of arrest, type of offense, and the like. Subsequently, however, staff from the District Attorney's office looked up each youth and indicated if there was a juvenile record – no other information was provided nor did the DA's office appear to review the adult arrest records. The lack of access to the detailed records and to the adult charges makes it difficult to assert that the findings reported here are reliable.

[28] For more on this topic, see https://www.nij.gov/topics/crime/Pages/delinquency-to-adult-offending.aspx

In order to maximize the data provided, a series of rules were implemented to try to address these issues.

First, all youth with a juvenile arrest were coded as recidivists. Second, among those that had *only* an adult arrest, if they were 17 or older at the time of program participation, they were coded as recidivists. Those who were from ages 13 to 16 (or without an age) with *only* an adult arrest reported, yet that arrest occurred within 300 days of the program participation, they were classified as recidivists. Finally, among those no date of arrest but the record contained an indicator of a recidivism time period (e.g., within 1 year) these were coded as recidivists. All others were non-recidivists.

Second, in order to report out on recidivism at 1, 2, and 3-year post-program participation, we combined a series of variables to create approximate dates of arrest. The actual date of arrest for the 274 cases with dates were chosen. Then for the 220 cases with data provided in the monthly DA spreadsheets, the date was the 15th of the month of the date of the spreadsheet. Then the number of days from arrest was calculated by subtracting the program participation date from this date. Among the remaining 132 with a recidivism period indicator (e.g., 0 = re-arrest within the first year, 1=re-arrest within 1 to 2 years etc.,) we used the midpoint of the period so that those arrested within the first year were coded as having a first arrest within 182 days; arrested within 1 to 2 years, days to arrest was 548; and those from 2 to 3 years, the days to arrest was 821. A date of arrest was approximated by adding the midpoint days to the program attendance date. Finally, for recidivists without any other information with which to approximate a date, we used the number of days since they were in the program.[29] Admittedly, this is a more conservative measure (e.g., likely overestimates the length of time until the first arrest) but it does allow us to use all the data in exploring this outcome.

Nonetheless, it is possible that these rules still fail to accurately capture recidivism. Given these limitations, we recommend that the following recidivism results be viewed cautiously.

Overall and by Year

Looking at the data overall and by year, we note the re-arrest rates in Table 10 indicate that overall, 82% of the 4,038 youth were not rearrested following participation in the workshop. By year, the highest number of youth rearrested were from the 2013 cohort – 152 (or 31%) were arrested. Very few of the youth in 2017 were rearrested – 6 of 469 (or 1%) (likely a consequence of the limited time at risk for re-arrest).

Looking at re-arrest within 1, 2, 3 and more than 3-years post-program participation, we note that overall among the 727 rearrested:

- 274 (38%) were rearrested within the first year after participation in the program;
- 218 (30%) were rearrested within the second year;
- 116 (16%) were rearrested within the third year; and
- 119 (16%) were rearrested three or more years post program participation.

[29] Calculated by subtracting the date of program attendance from December 31, 2017. The average number of days since participation was 1,887 (over 5 years) and ranged from 15 to 3,645 days).

Reviewing this by year of participation, trends vary widely across the time periods. For instance, in 2013 within the first year, only 5% were rearrested, and the bulk of those rearrested (60%) was 3 or more years post participation. In contrast, in 2012, 73% of the 52 recidivists were rearrested within the first year, with 11% after 3 or more years.

Three important notes on these trends. First, these variations across the years could reflect the change in the data collection process from GSMP staff to the DA reporting the data, and/or could reflect different frequencies of when these records were looked up and recorded (e.g., records reviewed annually versus monthly). Second, this may be an artifact of using approximate dates of first arrest. As noted above, 453 of 727 (62%) of the dates of arrest were approximated from other available data information. For example, for the 101 cases where we used the days since program participation as the days to first arrest may be influencing this trend (particularly as the switch from GSMP staff to DA data collection occurred in 2014 – 2013 may reflect a transition year). Finally, the year by year analysis includes small numbers. This is an issue because with larger sample sizes (usually around 100 observations) data tends to cluster around a midpoint into a bell-shaped curve.[30] With small samples, the results are more susceptible to the influence of outliers in the data (e.g., one or two youth with data that strongly differ from the other youth). Thus, while these time to re-arrest results are informative, they should not be overstated. In the limitations section, we discuss ways to improve the data collection process and research design to more clearly test the efficacy of this program.

[30] This refers to the Central Limit Theorem. See: https://en.wikipedia.org/wiki/Central_limit_theorem

Table 10: Recidivism Outcomes Overall and By Year N=4,038

Year		Rearrested Overall				Rearrested Within 1, 2, 3 or 3 or More Years								
	N[4]	No		Yes		N	Within 1 Year		Within 2 Years		Within 3 Years		More than 3 Years	
		Freq	%	Freq	%		Freq	%	Freq	%	Freq	%	Freq	%
Overall	**4,038**	**3,311**	**82%**	**727**	**18%**	***727***	**274**	**38%**	**218**	**30%**	**116**	**16%**	**119**	**16%**
2008	555	421	76%	134	24%	*134*	58	43%	53	40%	17	13%	6	5%
2009	508	431	85%	77	15%	*77*	52	67%	20	26%	5	6%	0	0%
2010	396	318	80%	78	20%	*78*	49	63%	26	33%	0	0%	3	4%
2011	159	127	80%	32	20%	*32*	20	63%	11	34%	1	3%	0	0%
2012	431	379	88%	52	12%	*52*	38	73%	8	15%	0	0%	6	11%
2013	492	340	69%	152	31%	*152*	7	5%	1	<1%	52	34%	92	60%
2014	190	148	78%	42	22%	*42*	2	5%	14	33%	14	33%	12	28%
2015	422	312	74%	110	26%	*110*	37	34%	46	42%	27	25%	*N/A*	*N/A*
2016	416	372	89%	44	11%	*44*	5	11%	39	89%	*N/A*	*N/A*	*N/A*	*N/A*
2017	469	463	99%	6	1%	*6*	6	100%	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*

*Among the 727 youth with a date of arrest or an approximated date of arrest.
N/A = Not enough time passed to assess recidivism for this time period.

Logistic Regression Analysis

The final step in the recidivism analysis is to conduct a regression analysis to predict the outcome, while accounting for information contained in other variables which could explain that outcome. For example, male offenders are more likely to recidivate, thus one would want to "control" for gender in the analytic model. For the recidivism analysis, we use logistic regression which is better suited to binary outcomes (arrest/no arrest) than the OLS regression used to assess the constructs.

Table 11 provides 3 separate models[31] looking at the likelihood of arrest among GSMP participants. The first model includes only age and gender (males=1, female=0). Odds ratios that are above 1 have a positive relationship with the outcome and below 1 have a negative (or opposite) relationship to the variable of interest (in this case arrest). In this first model, male participants have a higher likelihood of recidivism. To ease interpretation, we converted the odds ratios to predicted probabilities[32]. For Model 1, males are 9% more likely to be arrested than females, even after accounting for the age of the participant.

In Model 2, we added offense data – including whether the youth had committed a person offense, drug offense, weapon offense or auto theft. Gender is still significant (with these additional variables in the model, males are now 7% more likely to be arrested); as are those charged with drug, weapons, and auto theft. Youth who participated in the auto theft module are 10% more likely to be arrested and those charged with a drug offense are 6% more likely. However, those charged with a weapons crime are 6% *less* likely to be arrested.

Finally, in Model 3 key program variables were included in the model including if the referral to GSMP was from the court, the number of days since the youth participated in the program[33] and the pre-test scores from Construct 4 – Agency/Potential. Here we see that the gender and offense type variables remain statistically significant, even after controlling for these program factors. In fact, those who are referred from the court have a higher predicted probability of arrest of 4% (compared to those referred from another source); and while the number of days is statistically significant (so it is important to include it in the model), there is no independent impact of the number of days (as indicated by a 0% predicted probability of arrest). We also included the pre-test Construct 4 Agency/Potential score. Arguably, those with a higher level of agency and potential prior to exposure to GSMP would be less likely to be arrested irrespective of the impact of the program. Including this measure allows us to try to account for this factor. This was a statistically significant factor in that those lower scores on the pre-test C4 are 7% *more* likely to be arrested and those with higher C4 pre-test scores.[34]

[31] We conducted the logistic regression analysis including a number of variables including race, ethnicity, cohort 2, the 4 constructs, and the difference from pre- to post-test scores in the 4 constructs. Ultimately, the variables in Table 11 were statistically significant or if not significant, they were retained as they were theoretically important.
[32] Probabilities were calculated based on output values, and the calculation worksheets are provided in Appendix H.
[33] Calculated by subtracting the date of program attendance from December 31, 2017. This variable was included to account for the time at risk for recidivism.
[34] We ran this model only including C4 Post-Test score and the predicted probability was similar.

Table 11: Logistic Regression -- Predicting Recidivism

Logistic Odds Ratios[#] and z Statistic	Model 1 – Demos	Model 2 – Demos + Offense Type	Model 3 – Full Model
Independent variables			
Age of Youth	1.01 (.74)	0.97 (-1.05)	0.95 (-1.75)
Male (Yes=1; No=0)	1.93 (6.78)***	1.65 (4.82)***	1.64 (4.19)***
Person (Yes=1; 0=No)		0.98 (-.19)	1.03 (.24)
Drug (Yes=1; 0=No)		1.45 (2.63)**	1.58 (2.80)**
Weapon (Yes=1; 0=No)		0.61 (-2.98)**	.61 (-2.67)**
Auto Theft Module (Yes=1; 0=No)		1.85 (4.18)***	1.98 (4.20)***
Court Referral (Yes=1; 0=No)			1.25 (1.98)*
# Days Since Attended Program			1.00 (5.41)***
C4: Agency/Potential Pre-Test Average Scale Score			0.74 (-3.73)***
Constant	.11 (6.08)***	.22 (-3.78)***	0.57 (-1.06)
Observations	3,814	3,644	2,761
Pseudo R-Square	.014	.025	.045
Predicted Probability	.178	.173	.179
Log Likelihood	-1788.64	-1677.04	-1290.83

Statistically significant levels *p <.05, **p<.01, ***p <.001

[#]Odds Ratios with values above 1 indicate a positive association (or higher odds of the outcome occurring), values below 1 indicate a negative association (or lower odds of the outcome occurring).

Auto Theft and Retail Theft Module Specific Reports

Auto Theft Module Participants and Outcomes

The conflict resolution workshop includes a module focused on auto theft. Participants attending the workshop who had an auto offense are examined separately in this section of the report.

Demographics

presents the descriptive statistics for the 349 youth with an auto offense who participated in the module and show the age of participants ranged from 11 to 19 years with a mean of 15.9 years. A majority of the participants were male (93%). 70% of participants were Black/African American. Sixteen percent of youth identified as Hispanic. Fifty-nine percent of participants were referred by a Youth Aid Panel (YAP) and 36% were court ordered.

Table 12: Demographics of Youth with an Auto Offense Referred to GSMP (2008-2017)

	N[4]	Freq.	Percent	Range	Mean (SD)[4]
Demographics					
Age	342			11 to 19	15.92 (1.42)
Race	344				
Black/African American		241	70%		
White/Caucasian		40	12%		
Other		63	18%		
Ethnicity	344				
Hispanic		57	16%		
Sex	345				
Male		319	93%		
Female		26	7%		
Referral Type	331				
Court Ordered		125	36%		
YAP		205	59%		
Other		1	0.3%		

Pre-Post Test Results

In order to examine how the auto theft module may have impacted youth perceptions, a paired samples t-test was conducted. Among the 349 youth referred to JODP with an auto offense, data were provided for 330 youth who completed a pre-test, a post-test, or *both* pre-and post-test. Among those 330, 98% (323) of youth had pre-test survey data and 93% (306) had post-survey data. 299 individuals who committed an auto theft offense completed both a pre and post-test survey.[35] Table 13 presents the results and shows significant changes in mean responses for 11 of the 23 survey questions.

[35] An analysis of attrition indicated no statistically significant differences between those who completed the post-test

Measures with a statistically significant increase (meaning participants were (on average) ***more*** likely to agree with the statement post-workshop), included:

- Fighting usually solves a problem.
- I know what things people do or say that trigger my anger.
- When I'm angry I feel I must act on my anger right away.
- When people make me angry I feel I am justified in acting out against them.
- I feel that I should be in this program.
- I think this workshop will be helpful to me.
- I believe that I will learn something from the program that will benefit me.

Measures that saw a statistically significant decrease, (meaning participants were (on average) ***less*** likely to agree with the statement), included:

- I understand coming to this workshop is one of the consequences of my actions.
- I am willing to take responsibility for my actions that led to my arrest.
- I understand the consequences that may happen to me if I commit another crime.
- I understand how my actions may affect others (family, friends, others) if I commit another crime.

These results indicate overall positive effects of the workshop, but also indicate some potential issues. While measures with significant increases indicate that participants have positive feelings towards the program, attitudinal changes, specifically toward physical violence (e.g., *fighting usually solves a problem)* and anger (e.g., *When people make me angry I feel I am justified in acting out against them*), are troubling. The measures with significant decreases indicate a potential need for greater emphasis on consequences and taking responsibility for actions in the workshop.

There were no significant differences by demographic factors among youth who attended the auto theft module over this 10-year period. However, it is possible that the differences we are finding in the pre- to post-test related to emphasis on consequences and taking responsibility for actions may be the result of different instructors over this course of time or other factors related to program delivery that we cannot account in these data. However, in an attempt to understand the findings, we explored a sub-sample of auto-theft cases and found some interesting results. Given the small number of cases in each year with data available to assess (e.g., of the 349 cases, 299 youth with both pre-and post-test data, with most of those missing from 2008 and 2009 workshops), we reanalyzed[36] the pre-post survey data ***excluding*** 2008 and 2009 students. There were 229 youth in the 2010-2017 classes, and we found that the following measures ***are no longer statistically significant:***

- Fighting usually solves a problem.
- When I'm angry I feel I must act on my anger right away.
- I understand coming to this workshop is one of the consequences of my actions.

and those who did not.

[36] Not shown but available upon request

Further, with this smaller group of workshop participants, the measure of "*It's possible for me to think about the consequences of my behavior before I act on my feelings*" shows a statistically significant increase.

Table 13: Auto Theft Offenders Pre-Post Test Outcomes N=299

Responses on a Scale of 1 to 5, where 1=strongly disagree to 5=strongly agree	**N[4]**	**Pre-Test Mean**	**Post-Test Mean**	**Diff**
Q1. It is okay to walk away from a fight whether or not you think you would win.	295	3.79	3.79	0.00
Q2. Fighting usually solves a problem.	295	2.14	2.28	0.14*
Q3. I can control how I am feeling at any given time.	293	3.94	3.95	0.01
Q4. I know what things people do or say that trigger my anger.	294	3.65	3.93	0.28***
Q5. When I'm angry I feel I must act on my anger right away.	296	2.32	2.51	0.19**
Q6. Listening to the person you are angry with does not do any good.	290	3.09	3.10	0.01
Q7. When people make me angry I feel I am justified in acting out against them.	294	2.35	2.61	0.26***
Q8. It's possible for me to think about the consequences of my behavior before I act on my feelings.	295	3.88	3.99	0.11
Q9. I can choose different ways of reacting to someone when they make me mad.	287	3.95	3.91	-0.04
Q10. It's possible for people to see the same situation in different ways.	283	3.92	3.86	-0.06
Q11. It's important for me to understand where the other person is coming from before I do something in a conflict.	285	3.91	3.85	-0.06
Q12. What I want/need are more important than what the other wants/needs if there is a problem.	270	2.87	2.94	0.07
Q13. I feel I should be in this program.	278	2.77	3.36	0.59***
Q14. I understand why I am in this program/ I understand why I was in this program. (post)	275	4.09	4.11	0.02
Q15. I think this workshop will be helpful to me/ This workshop helped me. (post)	276	3.68	4.13	0.45***
Q16. I believe I will learn something that will benefit me/I learned something that will benefit me. (post)	281	3.92	4.22	0.30***
Q17. I am responsible for the actions that brought me into this program.	280	4.13	4.05	-0.08
Q18. I understand coming to this workshop is one of the consequences of my actions.	281	4.25	4.10	-0.15*

Responses on a Scale of 1 to 5, where 1=strongly disagree to 5=strongly agree	N[4]	Pre-Test Mean	Post-Test Mean	Diff
Q19. I see how my actions have affected my relationships with family, friends, and others.	283	3.85	3.87	0.02
Q20. I am willing to take responsibility for my actions that led to my arrest.	279	4.25	4.12	-0.13*
Q21. I realize how serious my crime could have been.	279	4.20	4.19	-0.01
Q22. I understand the consequences that may happen to me if I commit another crime.	281	4.40	4.26	-0.14*
Q23. I understand how my actions may affect others (family, friends, others) if I commit another crime.	280	4.34	4.16	-0.18**

Statistically significant levels *P<.05,**P<.01, ***P<.001

Recidivism Analysis

As noted in the methodology section, GSMP provided spreadsheets with re-arrest data for the youth charged with auto theft who attended the workshop from the period of 2008 to 2017. In total there were 349 auto theft module attendees in the period from 2008 to 2017.

Looking at the data overall and by year, we note the re-arrest rates in Table 14 indicate that overall, 70% of the 349 youth were not rearrested following participation in the workshop.[37] By year, the highest number of youth rearrested were from the 2010 cohort – 20 (or 53%) were arrested. Of those who attended the auto theft module, none of the youth in 2017 were rearrested (likely a consequence of the limited time at risk for re-arrest).

Similar to the overall recidivism study – however, there are too few youth rearrested to break the data down by year -- we were able to look at the length of time to re-arrest. Overall, among the 106 youth who participated in the Auto Theft module and recidivated:

- 48 (45%) were rearrested within the first year;
- 42 (40%) were rearrested within the second year;
- 12 (11%) were rearrested within the third year; and
- 4 (4%) were rearrested three or more years post program participation.

[37] In a prior report focused on Auto Theft from 2013 to 2017, we also reported if the youth were rearrested for auto theft or some other crime. However, looking at the full 10-year period, too many cases were missing the recidivism offense to reliably report this information.

Table 14: Auto Theft Offenders Recidivism Outcomes Overall and By Year N=349

Year	Rearrested?				
	N[4]	Yes		No	
		Freq.	Percent	Freq.	Percent
Overall	**349**	**106**	**30%**	**243**	**70%**
2008	82	32	39%	50	61%
2009	33	12	36%	21	64%
2010	38	20	53%	18	47%
2011	17	6	35%	11	65%
2012	26	3	12%	23	88%
2013	18	5	28%	13	72%
2014	13	2	15%	11	85%
2015	46	20	43%	26	57%
2016	36	6	17%	30	83%
2017	40	0	0%	40	100%

Retail Theft Module Participants and Outcomes

The conflict resolution workshop also includes a module focused on retail theft. Participants attending the workshop who had a retail theft offense are examined separately in this section of the report.

Demographics

Table 15 displays the demographics for youth with a retail offense who participated in the module. The age of participants ranged from 10 to 18 years with a mean of 15.4 years. In contrast to the makeup of youth in other areas of the program, the majority of the youth with a retail offense that participated in the module were female (70%). 72% of participants were Black/African American. Eleven percent of youth identified as Hispanic. Eighty-four percent of participants were referred by a Youth Aid Panel (YAP) and 10% were court ordered.

Table 15: Demographics of Youth with Retail Theft Offense (2008-2017)

	N[4]	Freq.	Percent	Range	Mean (SD)[4]
Demographics					
Age	418			10 to 18	15.42 (1.61)
Race	425				
Black/African American		311	72%		
White/Caucasian		49	11%		
Other		65	15%		
Ethnicity	425				
Hispanic		45	11%		
Sex	426				
Male		126	29%		
Female		300	70%		
Referral Type	406				
Court Ordered		44	10%		
Community Referral		2	0.5%		
YAP		359	84%		
Other		1	0.2%		

Pre-Post Test Results

In order to examine how the retail theft module may have impacted youth perceptions, a paired samples t-test was once again conducted. Among the 430 youth referred to JODP with a retail theft offense, data were provided for 367 youth who completed a pre-test, a post-test, or *both* pre-and post-test. Among those 367, 98% (360) of youth had pre-test survey data and 98% (360) had post-survey data. 353 individuals who committed a retail theft offense completed both a pre

and post-test survey.[38] Table 16 presents the results and shows significant changes in mean responses for 11 of the 23 survey questions.

Measures with a statistically significant increase (meaning participants were (on average) ***more*** likely to agree with the statement post-workshop), included:

- It is okay to walk away from a fight whether or not you think you would win.
- I know what things people do or say that trigger my anger.
- It's possible for me to think about the consequences of my behavior before I act on my feelings.
- I can choose different ways of reacting to someone when they make me mad.
- It's possible for people to see the same situation in different ways.
- I feel I should be in this program.
- I understand why I am in this program/I understand why I was in this program. (post)
- I think this workshop will be helpful to me/This workshop helped me. (post)
- I believe I will learn something that will benefit me/I learned something that will benefit me. (post)

Measures that saw a statistically significant decrease, (meaning participants were (on average) ***less*** likely to agree with the statement), included:

- I understand the consequences that may happen to me if I commit another crime.
- I understand how my actions may affect others (family, friends, others) if I commit another crime.

These results largely indicate positive results of the program. In fact, all statistically significant increases were the desired result. However, the two questions that saw a statistically significant decrease may indicate a need for greater emphasis on consequences of actions, particularly in relation to effects on others. This theme carried through all participants and both the auto theft and retail theft subgroups.

Table 16: Retail Theft Offenders Pre-Post Test Outcomes N=353

Responses on a Scale of 1 to 5, where 1=strongly disagree to 5=strongly agree	N[4]	Pre-Test Mean	Post-Test Mean	Diff
Q1. It is okay to walk away from a fight whether or not you think you would win.	351	3.92	4.05	0.13*
Q2. Fighting usually solves a problem.	351	2.05	2.10	0.05
Q3. I can control how I am feeling at any given time.	346	3.63	3.63	0.00
Q4. I know what things people do or say that trigger my anger.	347	3.74	4.07	0.33***

[38] An analysis of attrition indicated no statistically significant differences between those who completed the post-test and those who did not.

Responses on a Scale of 1 to 5, where 1=strongly disagree to 5=strongly agree	N[4]	Pre-Test Mean	Post-Test Mean	Diff
Q5. When I'm angry I feel I must act on my anger right away.	351	2.37	2.33	-0.04
Q6. Listening to the person you are angry with does not do any good.	352	3.07	2.93	-0.14
Q7. When people make me angry I feel I am justified in acting out against them.	346	2.33	2.42	0.09
Q8. It's possible for me to think about the consequences of my behavior before I act on my feelings.	353	3.86	4.03	0.17**
Q9. I can choose different ways of reacting to someone when they make me mad.	344	3.90	4.05	0.15**
Q10. It's possible for people to see the same situation in different ways.	343	3.90	4.12	0.22***
Q11. It's important for me to understand where the other person is coming from before I do something in a conflict.	342	3.87	3.98	0.11
Q12. What I want/need are more important than what the other wants/needs if there is a problem.	326	2.79	2.81	0.02
Q13. I feel I should be in this program.	334	3.13	3.73	0.60***
Q14. I understand why I am in this program/ I understand why I was in this program. (post)	333	4.30	4.41	0.11*
Q15. I think this workshop will be helpful to me/ This workshop helped me. (post)	331	3.92	4.36	0.44***
Q16. I believe I will learn something that will benefit me/I learned something that will benefit me. (post)	335	4.13	4.35	0.22***
Q17. I am responsible for the actions that brought me into this program.	338	4.48	4.43	-0.05
Q18. I understand coming to this workshop is one of the consequences of my actions.	335	4.40	4.44	0.04
Q19. I see how my actions have affected my relationships with family, friends, and others.	338	4.13	4.20	0.07
Q20. I am willing to take responsibility for my actions that led to my arrest.	336	4.51	4.46	-0.05
Q21. I realize how serious my crime could have been.	337	4.43	4.44	0.01
Q22. I understand the consequences that may happen to me if I commit another crime.	338	4.62	4.52	-0.10*
Q23. I understand how my actions may affect others (family, friends, others) if I commit another crime.	337	4.56	4.46	-0.10**

Statistically significant levels *P<.05,**P<.01, ***P<.001

Recidivism Analysis

As noted in the methodology section, GSMP provided spreadsheets with re-arrest data for the youth charged with retail theft who attended the workshop from the period of 2008 to 2017. In total there were 430 retail theft module attendees in the period from 2008 to 2017.

Looking at the data overall and by year, we note the re-arrest rates in Table 17. Overall, 86% of the 430 youth were not rearrested following participation in the workshop. By year, the highest number of youth rearrested were from the 2015 cohort – 13 (or 33%) were arrested. No one who attended the retail theft module in 2016 was rearrested, and only 1 youth was rearrested in 2017.

Similar to the overall recidivism study, while there are too few youth rearrested to break the data down by year, we were able to look at the time fame of when participants were re-arrested. Overall, among the 61 youth who participated in the Retail Theft module and recidivated:

- 21 (34%) were rearrested in the first year;
- 20 (33%) were rearrested in the second year;
- 12 (20%) were rearrested in the third year; and
- 8 (13%) were rearrested three or more years post program participation.

Table 17: Retail Theft Offenders Recidivism Outcomes Overall and By Year N=430

Year	Rearrested?				
	N[4]	Yes		No	
		Freq.	Percent	Freq.	Percent
Overall	**430**	**61**	**14%**	**369**	**86%**
2008	87	14	16%	73	84%
2009	53	7	13%	46	87%
2010	54	12	22%	42	78%
2011	17	0	0%	17	100%
2012	54	6	11%	48	89%
2013	33	6	18%	27	82%
2014	18	2	11%	16	89%
2015	40	13	33%	27	68%
2016	31	0	0%	31	100%
2017	43	1	2%	42	98%

Limitations and Conclusion

Overall, the GSMP program is successful at shifting youth perspectives in key conceptual areas of actions and consequences (C1); anger and perspectives (C2); and program expectations (C3). Further, youth in the program have a recidivism rate of 18%. Re-arrest within 1, 2, 3 and more than 3-years post-program indicates that among the 727 rearrested:

- 274 (38%) were rearrested within the first year after participation in the program;
- 218 (30%) were rearrested within the second year;
- 116 (16%) were rearrested within the third year; and
- 119 (16%) were rearrested three or more years post program participation.

However, there are limitations to this study that should be considered when discussing these outcomes and for future development of the program.

First, GSMP might consider conducting a process evaluation. The purpose of conducting a process evaluation is to assess whether programs are implemented as they were intended. In turn, this allows you to evaluate the outcomes of your program with more confidence. Integrity of program implementation process is the key to the ability to assert that the program had the desired impact on the participants and knowledge gained from a process evaluation can be used to improve future program effectiveness.[39]

Second, given the limitations of the current strategy to obtain arrest data, GSMP should try to obtain official criminal history data to enhance the recidivism analysis. At a minimum, GSMP would benefit from not only receiving the indicator of a juvenile and/or an adult arrest, but the date of first arrest post program participation, offense description, and the date the data was collected. The data collection process should be also solidified including a schedule of when the data is to be gathered, at what point adult arrest records are examined, and establish a clear practice for when to check records for youth who reach the age of 18.

Finally, this program would benefit from the inclusion of a comparison group. This requires the identification of another group of youth that had no exposure to the program. While this study informs GSMP on the characteristics and patterns of their participants and survey outcomes, and provides predictors for re-arrest, it remains that without a comparison group we are unable to ascertain if this recidivism rate of 18% is lower, higher, or the same as any other group of similarly situated youth.

[39] For more on this topic, see http://choiceresearchassoc.com/documents/importance_of_conducting_pdes_2011.pdf

Appendix A: Demographics of Youth Referred to GSMP (2001-2017)

	N[4]	Freq.	Percent	Range	Mean (SD)[4]
Demographics					
Age	5825			7 to 26	15.08 (1.82)
Race	5803				
Black/African American		4324	75%		
White/Caucasian		546	9%		
Other		933	16%		
Ethnicity	5805				
Hispanic		768	13%		
Sex	5876				
Male		3733	64%		
Female		2143	37%		
Referral Type	5506				
Court Ordered		1149	21%		
Community Referral		37	0.7%		
YAP		4271	78%		
Other		49	0.9%		

Appendix B: Detailed Offense Type of Youth Referred to GSMP (2008-2017)

Offense Type	Detailed Offense Type	Freq.	Percent
Person	Simple Assault	889	22%
Property	Theft	731	18%
Weapon	Weapon Violation	521	13%
Drug	Drug Offense	459	11%
Property	Auto Theft	349	9%
Unknown	Unknown	264	7%
Person	Aggravated Assault	254	6%
Public Order	Terroristic Threats	121	3%
Public Order	Disorderly Conduct	93	2%
Person	Robbery	82	2%
Property	Trespassing	49	1%
Other	Community Referral	44	1%
Property	Vandalism	44	1%
Property	Burglary	24	1%
Other	Conspiracy	24	1%
Property	Stolen Property	13	<1%
Public Order	Resisting Arrest	10	<1%
Property	Criminal Mischief	9	<1%
Person	Attempted Murder	7	<1%
Person	Criminal Recklessness	6	<1%
Person	Harassment	6	<1%
Property	Arson	5	<1%
Person	Threatening	4	<1%
Public Order	Loitering	3	<1%
Public Order	Contempt of Court	2	<1%
Property	Animal Cruelty	1	<1%
Property	Breaking and Entering	1	<1%
Public Order	Driving Under the Influence	1	<1%
Public Order	False Police Report	1	<1%
Person	Intimidation	1	<1%
Public Order	Obstruction	1	<1%
Public Order	Panhandling	1	<1%
Person	Sexual Assault	1	<1%
Public Order	Soliciting	1	<1%
Public Order	Underage Drinking	1	<1%

Appendix C: Attrition Analysis N=3,515

	Post-Test Data N=3,429			No Post-Test Data N=86			Difference
	Mean	**SD**[5]	**N**[4]	**Mean**	**SD**	**N**	
Demographics							
Gender – Proportion Male	0.34	0.47	3324	0.31	0.47	84	-0.03
Age	15.23	1.82	3276	14.78	2.01	85	-0.45
Race – Proportion Black	0.75	0.43	3306	0.74	0.44	86	0.00
Ethnicity – Proportion Hispanic	0.14	0.35	3308	0.20	0.40	86	0.05
Survey Responses (Responses from 1 to 5, where 1=Strongly Disagree 5=Strongly Agree)							
Q1. It is okay to walk away from a fight whether or not you think you would win.	3.80	1.06	3406	3.77	1.02	84	-0.03
Q2. Fighting usually solves a problem.	2.17	0.99	3412	2.07	1.01	85	-0.10
Q3. I can control how I am feeling at any given time.	3.58	1.25	3395	3.63	1.12	83	0.05
Q4. I know what things people do or say that trigger my anger.	3.71	1.11	3403	3.64	1.06	85	-0.07
Q5. When I'm angry I feel I must act on my anger right away.	2.50	1.12	3409	2.31	0.98	85	-0.20
Q6. Listening to the person you are angry with does not do any good.	3.22	1.30	3393	3.17	1.22	84	-0.05
Q7. When people make me angry I feel I am justified in acting out against them.	2.50	1.07	3390	2.36	0.97	85	-0.13
Q8. It's possible for me to think about the consequences of my behavior before I act on my feelings.	3.71	1.06	3402	3.64	0.99	85	-0.08
Q9. I can choose different ways of reacting to someone when they make me mad.	3.85	0.96	3416	3.81	0.82	85	-0.04
Q10. It's possible for people to see the same situation in different ways.	3.85	0.94	3399	3.83	0.86	84	-0.02
Q11. It's important for me to understand where the other person is coming from before I do something in a conflict.	3.73	1.01	3410	3.75	0.83	85	0.02
Q12. What I want/need are more important than what the other wants/needs if there is a problem.	2.82	1.04	3221	2.58	1.02	84	-0.23
Q13. I feel that I should be in this program.	2.78	1.32	3304	2.90	1.27	84	0.13

	Post-Test Data N=3,429			No Post-Test Data N=86			Difference
	Mean	SD[5]	N[4]	Mean	SD	N	
Q14. I understand why I am in this program.	3.95	1.07	3304	4.04	0.80	84	0.09
Q15. I think this workshop will be helpful to me.	3.63	1.06	3299	3.73	0.88	85	0.10
Q16. I believe that I will learn something from the program that will benefit me.	3.86	0.96	3310	3.85	0.91	86	-0.01
Q17. I am responsible for the actions that brought me into this program.	4.16	0.98	3325	4.12	0.89	86	-0.04
Q18. I understand coming to this workshop is one of the consequences of my actions.	4.21	1.10	3326	4.20	0.75	85	-0.01
Q19. I see how my actions have affected my relationships with family, friends, and others.	3.79	1.18	3323	3.98	0.90	85	0.19
Q20. I am willing to take responsibility for my actions that led to my arrest.	4.19	0.89	3305	4.15	0.85	85	-0.04
Q21. I realize how serious my crime could have been.	4.11	0.99	3311	4.15	0.88	85	0.04
Q22. I understand the consequences that may happen to me if I commit another crime.	4.38	0.80	3312	4.34	0.73	85	-0.04
Q23. I understand how my actions may affect others (family, friends, others) if I commit another crime.	4.32	0.86	3310	4.33	0.73	85	0.01

Appendix D: Pre-Test Survey

Good Shepherd Mediation Program
Juvenile Offender Diversion Programs/Youth Delinquency and Violence Prevention Workshop
Survey Questions: Pre-Test

In order to participate in the Good Shepherd Mediation Program diversion/prevention workshop to which you were referred, we need you—the youth—to complete the following survey. Your answers will help us design the workshop to best meet your needs.

How well do these statements represent YOU? That is, how true are these statements about YOU? Please circle the answer that best represents what YOU think using a scale from 1 to 5.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. It is okay to walk away from a fight whether or not you think you would win.	1	2	3	4	5
2. Fighting usually solves a problem.	1	2	3	4	5
3. I can control how I am feeling at any given time.	1	2	3	4	5
4. I know what things people do or say that trigger my anger.	1	2	3	4	5
5. When I'm angry I feel I must act on my anger right away.	1	2	3	4	5
6. Listening to the person you are angry with does not do any good.	1	2	3	4	5
7. When people make me angry I feel I am justified in acting out against them.	1	2	3	4	5
8. It's possible for me to think about the consequences of my behavior before I act on my feelings.	1	2	3	4	5
9. I can choose different ways of reacting to someone when they make me mad.	1	2	3	4	5
10. It's possible for people to see the same situation in different ways.	1	2	3	4	5
11. It's important for me to understand where the other person is coming from before I do something in a conflict.	1	2	3	4	5

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
12. What I want/need are more important than what the other wants/needs if there is a problem	1	2	3	4	5
13. I feel that I should be in this program.	1	2	3	4	5
14. I understand why I am in this program.	1	2	3	4	5
15. I think this workshop will be helpful to me.	1	2	3	4	5
16. I believe that I will learn something from the program that will benefit me.	1	2	3	4	5
17. I am responsible for the actions that brought me into this program.	1	2	3	4	5
18. I understand coming to this workshop is one of the consequences of my actions.	1	2	3	4	5
19. I see how my actions have affected my relationships with family, friends and others.	1	2	3	4	5
20. I am willing to take responsibility for my actions that led to my arrest.	1	2	3	4	5
21. I realize how serious my crime could have been.	1	2	3	4	5
22. I understand the consequences that may happen to me if I commit another crime.	1	2	3	4	5
23. I understand how my actions may affect others (family, friends, others) if I commit another crime.	1	2	3	4	5

Thank you for completing this survey. YOU MUST bring this completed survey and your information sheet on the day of your Good Shepherd Mediation Program Workshop.

Appendix E: Post-Test Survey

Good Shepherd Mediation Program
Juvenile Offender Diversion Programs/Youth Delinquency and Violence Prevention Workshop
Survey Questions: Post-Test

Now that you have had an opportunity to participate in the workshop, we would like to see if you think the things you learned will be helpful to you. That is, we want to know if you think you will consider some of the things you have learned today to affect your future behavior and actions.

Please take a few minutes to read each of the following statements and let us know how you feel about them by circling the answer that is most true for you. Again, please be truthful when responding to these items. These items are to help YOU and learn more about YOU, in addition to help us continue in our efforts to meet the needs of future workshop participants.

Do you think you will use the ideas and information you learned today when you go home? Please think about how you might change the ways you behave in the future as a result of participating in this workshop when faced in similar situations that brought you here initially. Again, please circle the answer that best represents what YOU think using a scale from 1 to 5.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. It is okay to walk away from a fight whether or not you think you would win.	1	2	3	4	5
2. Fighting usually solves a problem.	1	2	3	4	5
3. I can control how I am feeling at any given time.	1	2	3	4	5
4. I know what things people do or say that trigger my anger.	1	2	3	4	5
5. When I'm angry I feel I must act on my anger right away.	1	2	3	4	5
6. Listening to the person you are angry with does not do any good.	1	2	3	4	5
7. When people make me angry I feel I am justified in acting out against them.	1	2	3	4	5
8. It's possible for me to think about the consequences of my behavior before I act on my feelings.	1	2	3	4	5

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
9. I can choose different ways of reacting to someone when they make me mad.	1	2	3	4	5
10. It's possible for people to see the same situation in different ways.	1	2	3	4	5
11. It's important for me to understand where the other person is coming from before I do something in a conflict.	1	2	3	4	5
12. What I want/need are more important than what the other wants/needs if there is a problem	1	2	3	4	5
13. I feel that I should have been in this program.	1	2	3	4	5
14. I understand why I was in this program.	1	2	3	4	5
15. I think this workshop was helpful to me.	1	2	3	4	5
16. I believe that I learned something from this program that will benefit me.	1	2	3	4	5
17. I am responsible for the actions that brought me into this program.	1	2	3	4	5
18. I understand the consequences of my actions that brought me to this workshop.	1	2	3	4	5
19. I see how my actions have affected my relationships with family, friends and others.	1	2	3	4	5
20. I am willing to take responsibility for my actions that led to my arrest.	1	2	3	4	5
21. I realize how serious my crime could have been.	1	2	3	4	5
22. I understand the consequences that may happen to me if I commit another crime.	1	2	3	4	5
23. I understand how my actions may affect others (family, friends, others) if I commit another crime.	1	2	3	4	5

Using your own words, please answer the following questions. If you need more space to write, please ask for another piece of paper.

1. What, if anything, did you learn from this class?

2. What was the *best* part of the class?

3. What would you change about the class (aside from how long it lasted)?

4. What did you think about the trainers?

5. Do you have any suggestion for the trainers?

6. Please add any comments that you would like to share here.

THANK YOU FOR YOUR PARTICIPATION AND GOOD LUCK!

Appendix F: Factor Loading Results

	Factor 1	Factor 2	Factor 3	Factor 4
Q1	0.27	0.39	0.25	**-0.43**
Q2*	0.17	**0.47**	0.20	-0.25
Q3	0.10	0.29	0.01	**-0.40**
Q4	0.23	-0.14	0.09	-0.28[40]
Q5*	0.03	**0.61**	0.01	-0.30
Q6*	0.04	0.34[40]	0.06	-0.10
Q7*	0.12	**0.73**	0.06	-0.28
Q8	0.28	0.25	0.14	**-0.62**
Q9	0.27	0.19	0.15	**-0.66**
Q10	0.31	0.14	0.12	**-0.53**
Q11	0.33	0.28	0.25	**-0.57**
Q12*	0.02	0.36[40]	-0.01	-0.09
Q13	0.30	-0.02	**0.62**	-0.03
Q14	0.54	0.07	**0.56**	-0.23
Q15	0.47	0.10	**0.91**	-0.23
Q16	0.52	0.13	**0.76**	-0.30
Q17	**0.70**	0.03	0.40	-0.25
Q18	**0.74**	0.04	0.40	-0.32
Q19	**0.61**	0.07	0.44	-0.25
Q20	**0.76**	0.07	0.37	-0.32
Q21	**0.71**	0.12	0.44	-0.33
Q22	**0.72**	0.08	0.31	-0.38
Q23	**0.71**	0.07	0.36	-0.38

*Note: Reverse coded so that the all the items in the scale are in the same direction (e.g., higher values = more positive response).

[40] While Q6*and Q12* loaded less than .40 in factor 2 when all variables were considered, factor analysis of the final individual scales which only included the relevant variables revealed that these items loaded above .40 (e.g., Q6* and Q12* both loaded at .51). In addition, while Q4 on factor 4 did not load above .40 in the final scale, the analysis revealed no substantive difference in the outcomes when Q4 was dropped from the scale. Thus, we retained all items in these scales despite the initial low factor loadings.

Appendix G: Pre-Post Outcomes for Latent Constructs

	N	Pre-Test Mean	Post-Test Mean	Diff
Construct 1-Actions/Consequences	3,280	4.16	4.16	.00
Construct 2-Anger/Perspective	3,440	3.36	3.34	-.02*
Construct 3-Program Expectations	3,284	3.55	3.99	.44***
Construct 4-Agency/Potential	3,438	3.74	3.88	.14***

Statistically significant levels *p<.05, **p<.01, ***p<.001

Appendix H: Conversion Odds Ratio to Probability

Model 1

ARREST	Odds Ratios	Reduction in odds	Change in Prob 0 to 1	Change into %			
Age	1.02	2%	0.002	0%	obs	3914	
Male	1.93	93%	0.090	9%	LL	-1258.21	
Constant	0.11				pseudo r2	0.0138	1%
					prvalue	0.1782	

Model 2

ARREST	Odds Ratios	Reduction in odds	Change in Prob 0 to 1	Change into %			
Age	0.97	-3%	-0.005	0%			
Male	1.65	65%	0.068	7%			
Person	0.98	-2%	-0.003	0%	obs	3644	
Drug	1.45	45%	0.058	6%	LL	-1677.04	
Weapon	0.61	-39%	-0.062	-6%	pseudo r2	0.0255	3%
AT Module	1.85	85%	0.103	10%	prvalue	0.1728	
Constant	0.22						

Model 3

ARREST	Odds Ratios	Reduction in odds	Change in Prob 0 to 1	Change into %			
Age	0.95	-5%	-0.011	-1%			
Male	1.64	64%	0.069	7%			
Person	1.03	3%	0.005	0%	obs	2695	
Drug	1.58	58%	0.075	8%	LL	-1258.21	
Weapon	0.61	-39%	-0.065	-6%	pseudo r2	0.045	5%
AT Module	1.98	98%	0.118	12%	prvalue	0.1791	
Court Referral	1.25	25%	0.034	3%			
Days from Program	1.00	0%	0.000	0%			
C4 Pre-Test	0.74	-26%	-0.070	-7%			
Constant	0.57						