EVALUATION OF THE STAFF TRAINING AIMED AT REDUCING REARREST (STARR) CURRICULUM IN THE MIDDLE DISTRICT OF FLORIDA

ADMINISTRATIVE DATA REPORT
PART 2: IMPACT OF STARR ON PROBATIONER OUTCOMES

for the

UNITED STATES PROBATION OFFICE
FOR THE MIDDLE DISTRICT OF FLORIDA

Jill Viglione, Ph.D.
University of Central Florida

Ryan M. Labrecque, Ph.D.
University of Central Florida

June 10, 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Overview

In April 2017, the United States Probation Office for the Middle District of Florida began implementation of the Staff Training Aimed at Reducing Rearrest (STARR) curriculum across its offices. This initiative is in line with the overall orientation of the agency to use evidence-based practices (EBPs) and programming scientifically proven to achieve positive outcomes. Further, implementation of STARR supports the agencies main goal to generate long-term positive change amongst its population served. In February 2018, the Middle District of Florida published a formal policy detailing their Charter for Excellence and an overview of the STARR training and official agency policy. Given their focus on outcomes and dedication to use of EBPs to reduce recidivism, the Middle District of Florida recognizes the importance of evaluating the effectiveness of the STARR training program across their agency.

The primary goal of the current evaluation is to evaluate the implementation of the STARR program in the Middle District of Florida. On March 1, 2019, a mandate was passed districtwide to require all trained probation officers to use a minimum of eight STARR skills per month. Officers were required to document all skill use in PACTS. In consultation with district executive staff and frontline staff, March 1, 2019 began the true date of STARR implementation. Prior to this date, even though users had been trained and were using the STARR skills, use was sporadic, limited, and may not have been logged into PACTS. Thus, the implementation period for this evaluation began March 1, 2019. The current report uses administrative data obtained from the district.

The key research questions associated with the administrative data were:

1. Do trained officers use STARR skills in line with Middle District of Florida policy?
   a. How does use of STARR skills vary across offices within the district?
   b. With which risk groups are officers most likely to use STARR skills with?
   c. What are the most frequently used STARR skills?

2. What is the impact of STARR training on probationer outcomes?
   a. Are probationers supervised by trained officers less likely to have a positive urinalysis?
   b. Are probationers supervised by trained officers less likely to receive a technical violation?
   c. Are probationers supervised by trained officers less likely to receive a revocation?
   d. Are probationers supervised by trained officers less likely to be arrested for a new criminal offense?

Method

Administrative data were obtained from the district's key data systems, SITS and PACTS. The SITS database houses data pertaining to submission of STARR audio recordings including user recordings and coach recordings, coach feedback forms, and record of booster training attendance. Data obtained from SITS included the type (e.g., which skill focused on) and length of booster trainings officers attended, the total number of user and coach recordings submitted, the total number of times users and coaches received feedback on their use of skills, and total number of times users and coaches did not receive feedback. The PACTS database houses all supervision data and officer record of use of STARR skills in their personal contacts with individuals under supervision. Data obtained from PACTS included total personal contacts, total personal contacts in which STARR was used, type of STARR skill used, as well as probationer records and demographic information for all individuals supervised on federal post-release probation in MDFL between March 1, 2019 and August 31, 2019 (N = 4,902). Lastly, demographic data for each officer in the district including tenure, gender, and date received user/coach training was obtained from agency staff. This report is presented in two parts. Part I examines whether trained officers use STARR skills in line with MDFL policy while Part II examines the impact of STARR training on probationer outcomes.

Key Findings

Part I

1. **Do trained officers use STARR skills in line with Middle District of Florida policy?**
    a. The average reported monthly use of STARR skills increased from 5% in September 2018 to a high of 30% in March 2019
    b. After March 1, 2019, officers used STARR skills 17% more frequently
    c. Approximately 86% of officers used STARR skills a minimum of 8 times per month in March 2019
        i. From April 2019 to August 2019, this decreased to between 65% and 76%
    d. On average, coaches received 2 recordings to review per month
        i. Coaches completed reviews and provided feedback 88% of the time

2. **Does use of STARR skills vary across offices within the district?**
    a. Male officers used STARR on average 10% more frequently compared to female officers
    b. Officers with smaller caseloads reported a greater average use of STARR skills
    c. Officers with a tenure between 5-10 years reported the greatest average use of STARR skills
        i. However, these differences based on caseload and tenure were not statistically significant

3. **Does use of STARR skills vary across officers within the district?**
    a. An increase in use of STARR skills was seen across every office in the district
    b. The greatest average skill use pre- and post-policy change was seen in Cocoa, with a 23% increase

4. **With which risk groups are officers most likely to use STARR skills with?**
    a. The most frequent use of STARR skills occurred with low/moderate risk individuals, followed by moderate risk.
    b. The least frequent use of STARR sills occurred with the highest risk individuals.
    c. After the policy change, officers used STARR skills with moderate risk clients 1.55% more frequently and 1.7% less frequently with low risk individuals.

5. **What are the most frequently used STARR skills?**
    a. The most frequently used skills were Effective Use of Reinforcement, Role Clarification, Role Clarification Review, and Effective Use of Disapproval

b. The least frequently used skills were Problem Solving, Reviewing the Cognitive Model, and Applying the Cognitive Model.
c. The largest increase in use of skills was seen in Role Clarification Review (13% increase) and Role Clarification (14% decrease).
d. There was a 5% decrease in Effective Use of Punishment and 4% decrease in Effective Use of Disapproval
e. There was a 3% increase in Problem Solving and 2% increase in Reviewing the Cognitive Model

Part II

6. Are probationers supervised by trained officers less likely to have a positive urinalysis?

a. Probationers supervised by trained officers were 1.6% more likely to have a positive urinalysis during the 12-month follow-up period
b. High risk probationers were 12.1% less likely to have a positive urinalysis test

7. Are probationers supervised by trained officers less likely to receive a technical violation?

a. Probationers supervised by trained officers were 4.7% more likely to receive a technical violation during the 12-month follow-up period, but this finding was not statistically significant
b. High risk probationers supervised by trained officers had 7.2% fewer technical violations during the 12-month follow-up period, but this finding was not statistically significant
c. Moderate risk probationers supervised by trained officers were 4% more likely to receive a technical violation
d. Low/moderate risk probationers were 0.8% more likely to receive a technical violation

8. Are probationers supervised by trained officers less likely to receive a revocation?

a. Probationers supervised by trained officers were 2.2% less likely to receive a probation revocation during the 12-month follow-up period
b. High risk probationers supervised by a trained officer had 5.8% fewer revocations during the 12-month follow-up period, but this finding was not statistically significant
c. Low/moderate (4.2%) and low risk (2.2%) probationers supervised by a trained officer both had fewer revocations, although not statistically significant

9. Are probationers supervised by trained officers less likely to be arrested for a new criminal offense?

a. Probationers supervised by trained officers were 0.4% less likely to be rearrested during the 12-month follow-up period
b. High risk probationers had 6.5% fewer rearrests, but this finding was not statistically significant
c. Low/moderate probationers supervised by a trained officer had 3% fewer rearrests, although not statistically significant

BACKGROUND

Overview of Evidence-Based Practices

At present, the focus on use of evidence-based practices (EBPs) has led correctional agencies to implement practices aligned with core correctional practices scientifically proven to reduce recidivism. To promote EBPs, researchers developed principles of effective intervention, which outline specific strategies and tools correctional practitioners can implement to reduce recidivism and improve other outcomes (e.g., Andrews & Dowden, 2006; Cullen & Gendreau, 2000; Lipsey & Cullen, 2007; Smith, Gendreau & Swartz, 2009). These principles reflect best practices for correctional agencies, including:

- Risk and needs assessment practices
- Individualizing services to target dynamic risk factors
- Incorporating treatment planning
- Providing rewards and sanctions at a ratio of at least 4 to 1
- Providing an integrated approach for offenders with multiple needs (Andrews & Bonta, 2010; Gendreau, Little & Goggin, 1996; Smith, Gendreau & Swartz, 2009)

Increasingly gaining attention in the field of corrections is the risk-need-responsivity model (RNR), which combines an actuarial, managerial approach with a rehabilitative, clinical model of supervision. The RNR model outlines several principles designed to generate effective interventions for offender populations with the ultimate goals of improving treatment for offenders and reducing recidivism (Andrews & Bonta, 2010). The RNR model is based on three core principles:

- ***Risk Principle*** – Prioritize supervision and treatment resources for higher risk (to recidivate) offenders.
- ***Need Principle*** – Target criminogenic needs (i.e., dynamic risk factors directly related to recidivism).
- ***Responsivity Principle*** – Use cognitive-behavioral treatment methods and tailor programming to the motivation, learning styles, and strengths of the offender.

Adherence to all three core principles of the RNR model can significantly reduce recidivism by as much as 35% (Bonta & Andrews, 2007).

Despite evidence surrounding the effectiveness of the RNR model and core correctional practices, research also identifies the challenge of organizational change and successful implementation. Existing research documents the challenges associated with implementing and sustaining EBPs within justice agencies alongside existing organizational culture (e.g., Battalino, Beutler & Shani, 1996; Rudes, 2012; Viglione, Rudes, & Taxman, 2015; Viglione, 2017). One mechanism to promote and support change and increase adherence to the RNR model in correctional organizations is through formal training curriculums. There have been several attempts to integrate the principles of effective interventions into community supervision settings via specialized curriculums for probation and parole officers (Bonta et al., 2010; Latessa et al., 2013; Robinson et al., 2011; Taxman, Henderson, Young & Farrell, 2012). These curriculums attempt to translate the principles of the RNR model into concrete training programs and practices to increase the knowledge, understanding, and application of RNR principles into daily practice.

Research on organizational change within probation and parole organizations finds that POs trained specifically on the principles of RNR demonstrate:

- Better adherence to the RNR principles in practice
- More frequently use cognitive-behavioral techniques
- Supervise offenders who are more likely to have better outcomes (Bonta et al., 2011; Young, Farrell, & Taxman, 2012).

Training focused on the RNR principles is critical in skill development, as adherence to the RNR model requires important behavioral changes in POs and the individuals they supervise. Correctional curriculums, such as STARR, serve as a primary tool for community supervision agencies to generate and sustain change in the roles and behaviors of front-line staff. They offer comprehensive skill development, providing a framework for probation and parole agencies and staff to align with changing roles and expectations as a result of scientific evidence on effective practices.

STARR Overview

The Staff Training Aimed at Reducing Rearrest (STARR) is a correctional curriculum that promotes the use of core correctional skills by community supervision officers, focusing on client-practitioner interactions (Robinson, Lowenkamp, Holsinger, VanBenschoten, Alexandar, & Oleson, 2012). The STARR curriculum emphasizes the use of cognitive-behavioral supervision strategies to address dynamic risk factors during interactions with offenders (Robinson et al., 2012; Robinson, VanBenschoten, Alexander, & Lowenkamp, 2011). The STARR curriculum, which is based on the RNR model, emphasizes several key skills:

- Active listening
- Role clarification
- Effective use of authority
- Effective disapproval
- Effective reinforcement
- Effective punishment
- Problem solving skills
- Teaching, applying, and reviewing the cognitive model (Robinson et al., 2011)

Training for STARR includes a three-and-a-half-day classroom training session that includes information on the development and theory behind STARR, which emphasizes the RNR model. In addition, each of the key skills are demonstrated and officers are provided with opportunities to take part in exercises and practice each skill to receive feedback (Robinson et al., 2011, 2012). As part of the training, skill cards are used to provide strategies regarding specific actions and activities officers can do in order to deliver each of the skills taught in the curriculum. Skills are also taught through video examples and in-person demonstrations and practice. Officers also send in audiotaped interactions to demonstrate their understanding and use of skills as well as provide the opportunity for feedback on performance. Finally, four follow-up training sessions are held throughout the following year to provide additional training specifically emphasizing challenges that are recognized through the audiotape analysis (Robinson et al., 2011; 2012).

Research on STARR

Existing evaluations of the STARR curriculum report positive feedback. Probation officers who went through the training utilized reinforcement and disapproval more than untrained officers, and were more likely to discuss cognitions, peers and impulsivity with offenders (Robinson et al., 2011). A study of outcomes found that offenders supervised by officers trained in STARR had reduced failure rates, but when examining failure rates by risk level, there were no significant differences for higher-risk offenders (Robinson et al., 2011). In a second evaluation of the STARR curriculum, Robinson and colleagues (2012) found a similar trend in regard to offender outcomes. In addition, they found that while trained POs exhibited significantly greater use of core correctional skills, they still used those skills in fewer than 50 percent of their interactions with clients (Robinson et al., 2012). Lowenkamp and colleagues (2014) extended these earlier analyses to examine outcomes at 24-months, finding reduced failure rates for offenders supervised by trained officers, but little difference in the failure rates of high risk offenders pre- and post-training. The authors concluded the impacts of the STARR curriculum were most prominent for moderate risk offenders (Lowenkamp et al., 2014). However, failure rates for high risk offenders were reduced when their supervising officer was training in both STARR and motivational interviewing (MI) (Lowenkamp et al., 2014).

A recent study conducted by Hicks and colleagues (2020), examined the relationship between individual characteristics of probationers, drug test results, and federal supervision outcomes. Because all officers in the study had received STARR training, they controlled for whether trained officers were proficient in STARR in their analyses. Results indicated that drug tests taken by officers proficient in STARR were 1.7 times more likely to be positive than a drug test taken by someone supervised by a non-STARR proficient officer. There was no statistically significant impact of STARR proficiency on revocation for technical violations or drug violations. However, when probationers committed a new crime, those supervised by non-STARR proficient officers were almost twice as likely to receive a revocation compared to probationers supervised by a STARR proficient officer. Because the authors did not set out to specifically examine the relationship between STARR and supervision outcomes, this evidence is considered preliminary.

A recent article examined offender satisfaction of their interactions with their probation officer, specifically related to fairness, respect, and consistency (Alarid & Jones, 2018). Results indicated offenders reported positive perceptions of their supervising officers who had been trained in STARR, which increased over time. More specifically, offenders reported their officers gave them clear instructions and provided positive reinforcement, provided assistance developing a case plan and problem solving, felt their officer listened to them, believed they could be truthful to their officer without fear of revocation, viewed their officer as a role model, and believed their officer responded fairly to violations (Alarid & Jones, 2018).

Overall, the research on STARR suggests positive outcomes both in terms of reduced failure rates, reduced revocations for a new crime, improved use of evidence-based supervision skills, and offender satisfaction. However, research also identifies the challenge of successful implementation, particularly related to encouraging wide use of newly learned skills and implementation of skills with more challenging populations (i.e., high risk). Thus, while STARR is built on solid theory and existing evidence is promising, including a rating by Crime Solutions (crimesolutions.gov) as “promising”, evaluation of additional implementation efforts is critical to understand the effectiveness of the training program for specific agencies and their populations. Given the Middle District of Florida's emphasis on achieving long-term positive change and use of effective programming, the goal of this proposal is to recommend a thorough strategy for evaluating the use of EBPs in the District and the impact of the STARR curriculum on key agency outcomes.

EVALUATION METHODOLOGY

The primary goal of this evaluation is to evaluate the implementation of the STARR program in the Middle District of Florida. The primary research questions are:

Process Evaluation

1. What are probation staff attitudes and perceptions surrounding the STARR curriculum, training, and implementation process?
2. How do STARR users and coaches implement the key components of the STARR curriculum?
3. What are facilitators and barriers of STARR implementation?

Outcome Evaluation

4. Do trained officers use STARR skills in line with Middle District of Florida policy?
5. What is the impact of STARR training on probationer outcomes?

To address the first three research questions, organizational surveys and semi-structured interviews with all trained probation staff were conducted. The current report focuses on the results and recommendations related to research question 4, "Do trained officers use STARR skills in line with Middle District of Florida policy?" in Part I. PART II examines research question 5, "What is the impact of STARR training on probationer outcomes?".

Administrative Data

To examine this research question, data were obtained from the district's two key data management systems, SITS and PACTS. The SITS database houses data pertaining to submission of STARR audio recordings including user recordings and coach recordings, coach feedback forms, and record of booster training attendance. Data obtained from SITS included the type (e.g., which skill focused on) and length of booster trainings officers attended, the total number of user and coach recordings submitted, the total number of times users and coaches received feedback on their use of skills, and total number of times users and coaches did not receive feedback. Demographic data for each officer in the district including tenure, gender, and date received user/coach training was obtained from agency staff.

The PACTS database houses all supervision data and officer record of use of STARR skills in their personal contacts with individuals under supervision. Data obtained from PACTS included total personal contacts, total personal contacts in which STARR was used, type of STARR skill used, as well as probationer records and demographic information for all individuals supervised on federal post-release probation in MDFL between March 1, 2019 and August 31, 2019 (N = 4,902). There were 331 cases with one or more elements of missing probationer information. The final sample size for this portion of the study, therefore, includes 4,571 unique probationers.

PART I: DO TRAINED OFFICERS USE STARR SKILLS IN LINE WITH MIDDLE DISTRICT OF FLORIDA POLICY?

Analytic Plan

Because the focus of this component of the evaluation is to examine whether officers use STARR skills in line with district policy, a series of bivariate analyses was conducted to examine change in use of skills pre-policy change (September 1, 2018 to February 28, 2019) to post-policy change (March 1, 2019 to August 31, 2019). Descriptive statistics and bivariate analyses for all measures were reported for the entire district, with total use of skills also reported by office.

Sample

A total of 96 officers comprised the final sample for this evaluation. Officers included in the final dataset were primarily male (56%), white (64%), held a master's degree (48%), were on average 42 years old, and have been employed in MDFL for an average of 11 years. Most fulfilled non-supervisory roles (59%) and supervised an active caseload of 58 individuals. Of the 96 officers in the sample, 53% were trained in STARR at the time of data collection and 76% were trained as a coach. On average, users had been trained for an average of 22 months and coaches for an average of 21 months. At the time of data collection (9/01/2019), trained officers logged using at least one STARR skill in an average of 131 personal contacts. In terms of training, users and coaches attended an average of 1.65 boosters, spending an average of 341 minutes in booster training. The greatest amount of time spent in booster training was for coaching (480 minutes), followed by effective use of authority (258 minutes), and effective use of disapproval (238 minutes). The least amount of time spent in booster training was for supervisor boosters (30 minutes), role clarification (140 minutes), and problem solving (173 minutes).

For the analyses that measure change over time, the sample was reduced to reflect only those officers who received training and were active STARR users/coaches during at least one time point before the policy change. This resulted in sample of 43 eligible officers (21 users and 22 coaches).

Table 1. Probation Officer Characteristics (N = 96)

Demographic Variables	**Percent (*n*)**	**M**	**SD**	**Range**
Gender				
Male	56% (54)	-	-	-
Female	44% (42)	-	-	-
Race				
White	64% (61)	-	-	-
Nonwhite	25% (24)	-	-	-
Age	-	42	7.51	29-56
Position/rank				
PO	59% (57)	-	-	-
Supervisor	41% (39)	-	-	-
Caseload	-	58	55	0-350
Tenure+	-	11	8	1-28
Education				
Bachelor’s	35% (34)	-	-	-
Master’s	48% (46)	-	-	-
JD	2% (2)	-	-	-
STARR trained				
Yes	53% (51)	-	-	-
No	47% (45)	-	-	-
Coach trained				
Yes	24% (23)	-	-	-
No	76% (73)	-	-	-
Length of time trained+				
Users	-	22	8.76	4-42
Coaches	-	21	6.72	4-29
Total STARR Chronos!	-	131	85.51	0-327
Total Boosters Attended	-	1.65	2.55	0-8
Time Spent in Boosters^	-	341	535.68	0-1890
Effective Use of Disapproval	-	236	124.10	0-360
Effective Use of Reinforcement	-	182	92.03	0-420
Coaching	-	480	0	480-480
Supervisor	-	30	42.43	0-60
Problem Solving	-	173	76.16	0-300
Reviewing the Cognitive Model	-	200	108.17	0-360
Applying the Cognitive Model	-	192	107.33	0-240
Teaching the Cognitive Model	-	217	84.47	180-420
Role Clarification	-	140	72.66	0-180
Effective Use of Authority	-	258	91.87	180-480
Effective Use of Punishment	-	180	0	180-480
User Recordings Submitted	-	13.5	5	3-24
Coach Recordings Submitted	-	8.45	5	1-16
Users Received Review	-	11.48	4.41	3-21
User Reviews Not Received	-	1.6	2.89	0-14
Did Not Provide User Coaching	-	3.74	6.15	0-22
Total Coaching Recordings Provided	-	29.87	23.29	0-95

+Tenure in years; length of time trained in months; time spent in boosters in minutes

PART I: RESULTS

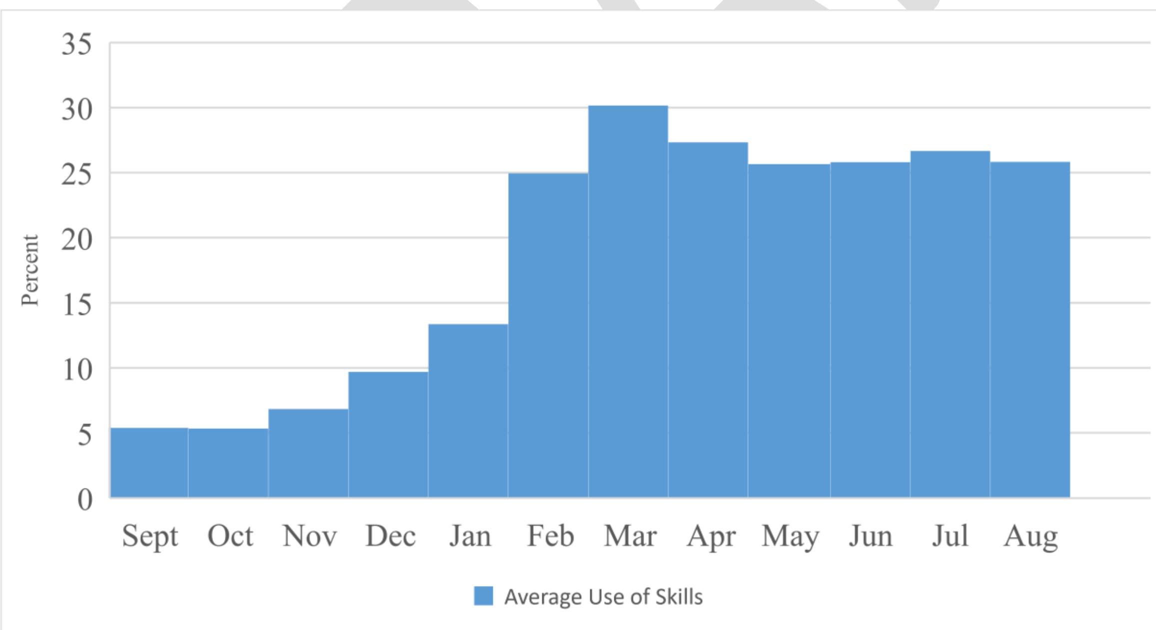
Do trained officers use STARR skills in line with Middle District of Florida policy?

Average Monthly Skill Use

First, the proportion of STARR skill use was calculated by creating four total variables per trained officer – the total number of personal contacts recorded six months before the policy change, the total number of personal contacts recorded in the six months after the policy change, the total recorded use of STARR skills before the policy change, and the total recorded use of STARR skills after the policy change. A proportion of STARR skill use was calculated for each office by dividing the total STARR skills used by the total personal contacts. In creating this proportion of STARR use of skills, it accounts for the total number of opportunities an officer had to implement the skills.

As seen in Figure 1, the average monthly use of STARR skills increased from November 2018 to January 2019. In September 2018, trained officers used STARR skills on average in 5% of their personal contacts per month. This average use increased to 25% of personal contacts in February 2019. In March 2019, directly following the policy change, the average use of skills was the highest with officers using STARR skills in approximately 30% of their personal contacts. From April to August 2019, the average use of skills leveled off to between 26% and 27%.

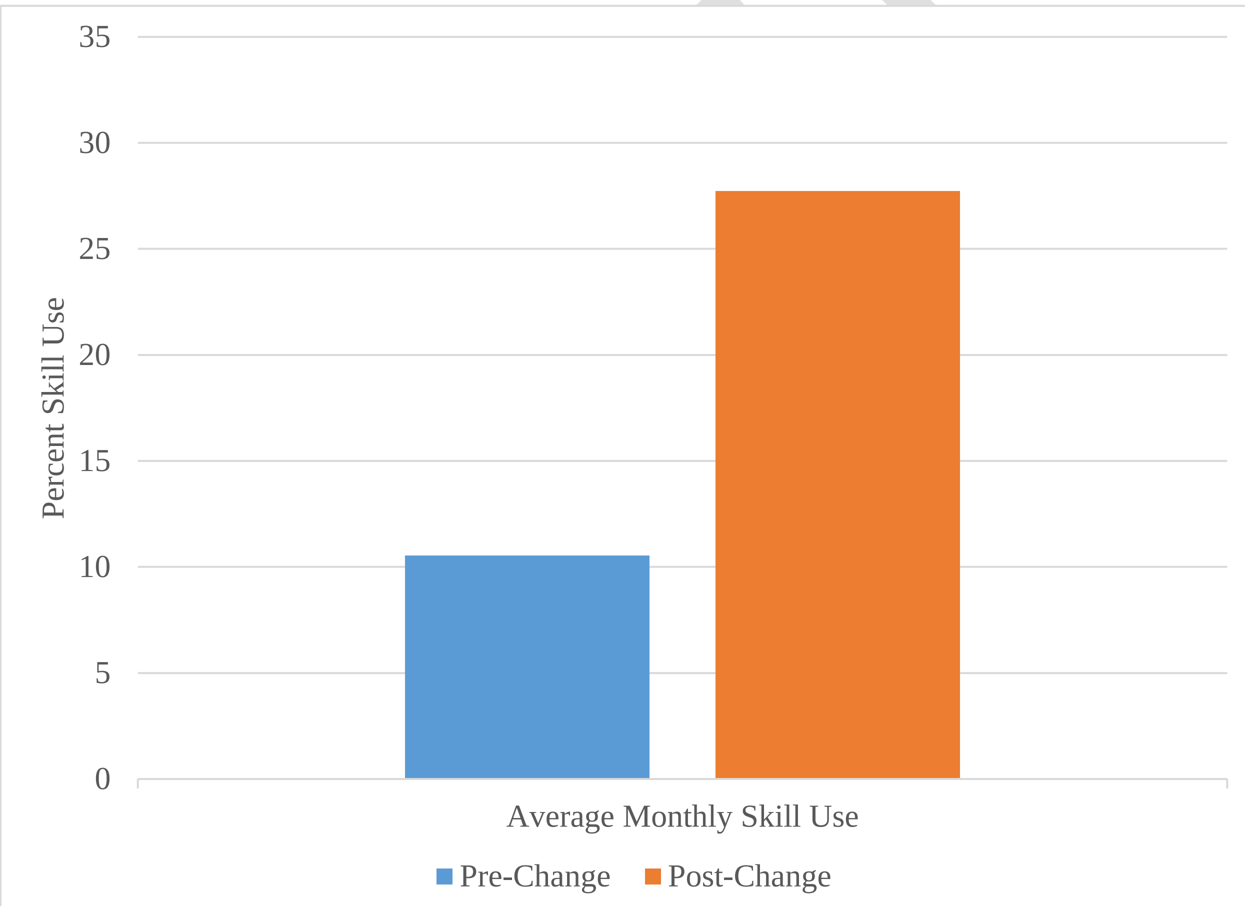
Figure 1. Average Monthly STARR Skill Use



Use of STARR Skills

To measure whether there was a statistically significant mean difference between the average skill use prior to the policy change enacted on March 1, 2019 and after the policy change, a paired samples t-test was used. Effect size was calculated using Cohen's *d* to measure whether the magnitude of the difference in means between two groups has practical significance (Cohen, 196) following the guidelines of .2 (small), .5 (medium), and .8 (large) (Cohen, 1988). Results indicated that trained officers used STARR skills at a higher proportion in their personal contacts after the policy change (an average of 27.72 times) compared to before the policy change (an average of 10.51 times), $t(44) = 10.39$, $p < .001$, $d = 1.55$. Post-policy change, the proportion of skill use increased by 17.21% (95% CI, 13.87 to 20.55) compared to pre-policy change, representing a large effect size.

Figure 2. Average Skill Use Across the Middle District of Florida



As seen in Figure 2, after the agency passed the policy requiring officers to use a minimum of eight STARR skills per month, trained officers used significantly more STARR skills in their personal contacts. Overall, officers used STARR skills 17% more frequently after March 1, 2019.

STARR Users

The policy that went into place on March 1, 2019 required all STARR users to use a minimum of eight STARR skills per month. Figure 3 reflects the total number and percentage of staff who adhered to this policy. As illustrated, the highest adherence rate was in March (86% of users) immediately following policy implementation.

Figure 3. Percentage of Users Who Submitted a Minimum of 8 Tapes Per Month

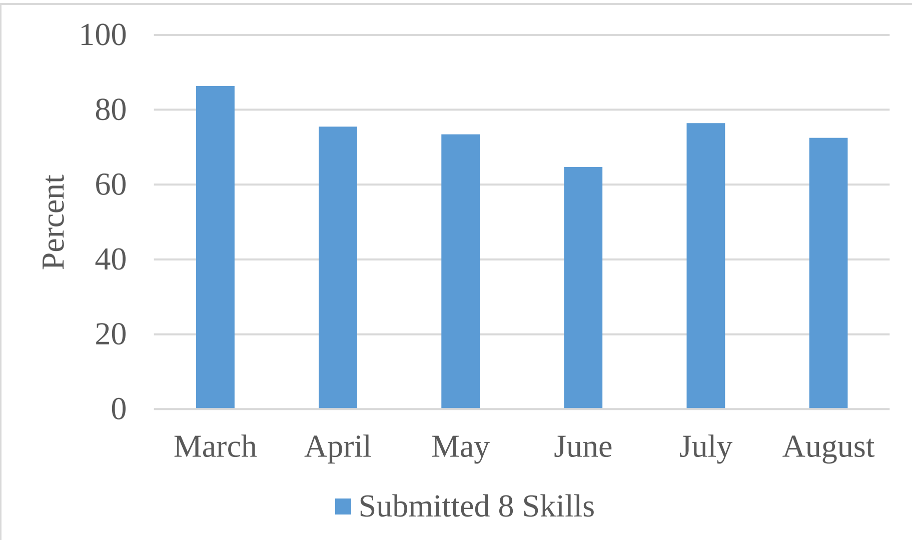
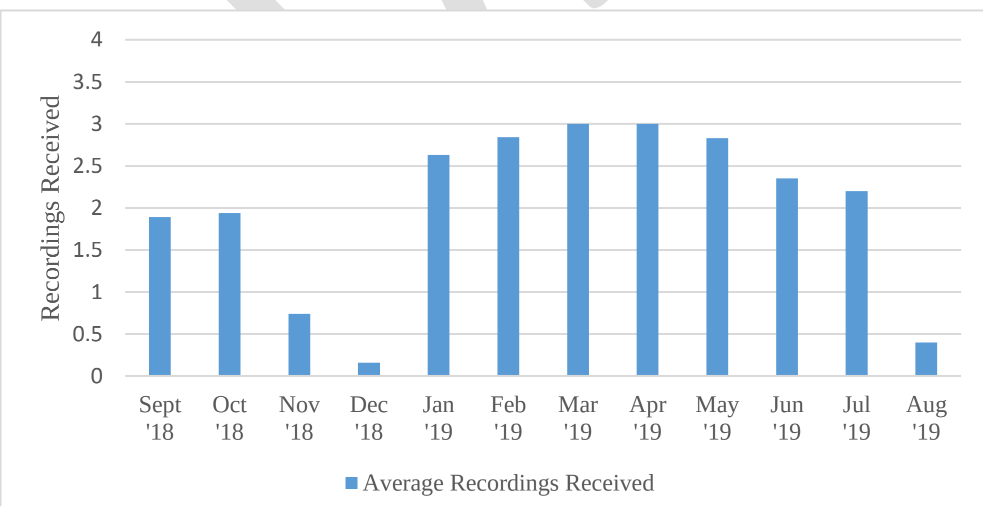


Figure 4 illustrates the average number of skill recordings coaches received to review during the study period. On average, coaches received 2 recordings to review per month. However, tapes were not distributed equally, with larger numbers of tapes to review on average in March and April 2019 and fewer tapes to review on average in November and December 2018 and August 2019. Coaches completed their reviews and provided feedback 88% of the time.

Figure 4. Average Skill Recordings Coaches Received

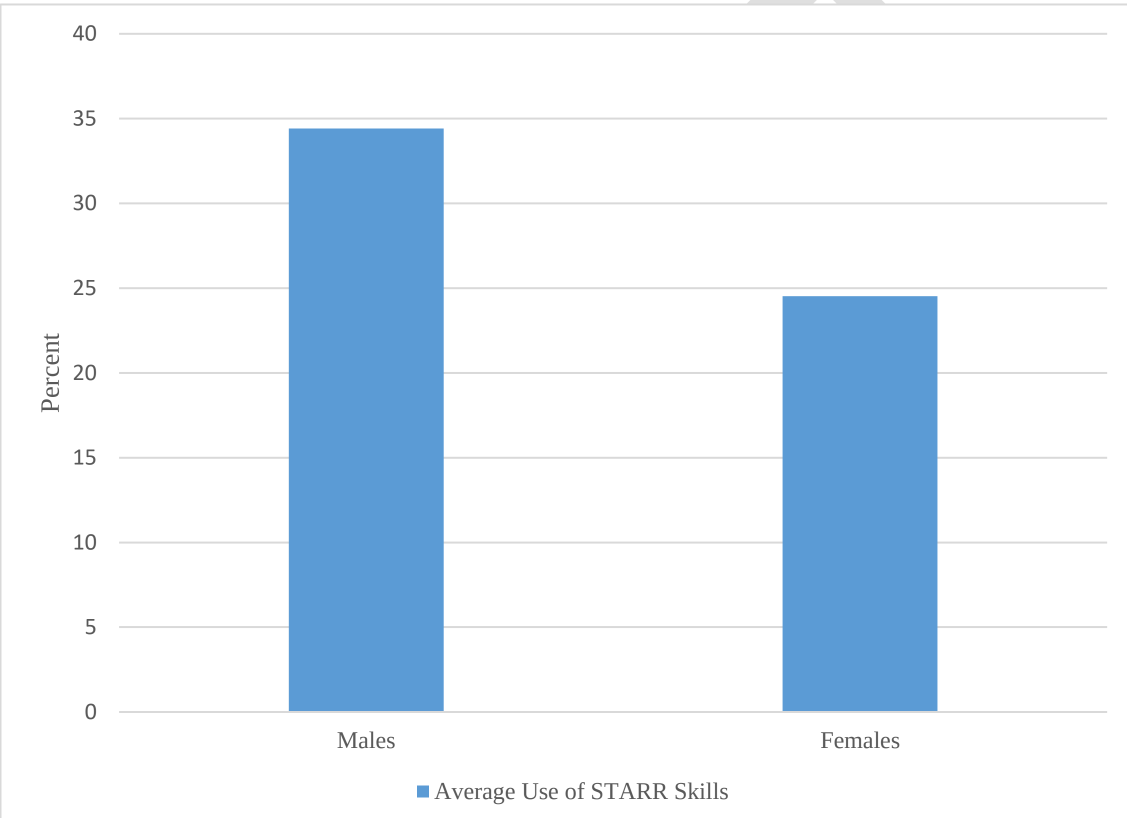


How does use of STARR skills vary across officers within the district?

Gender

An independent samples t-test was conducted to examine differences in use of skills based on gender. Of officers trained in STARR and an active user, 21 were males and 22 were females. The average use of skills between March 1, 2019 and August 31, 2019 was higher among male officers (an average of 34.42) compared to female officers (an average of 24.52), $t(4) = 2.93$, $p < 0.01$, $d = 2.56$. Average use of skills among male officers was 10% higher than average use of skills amongst female officers (95% CI, 3.03, 16.77), representing a large effect size.

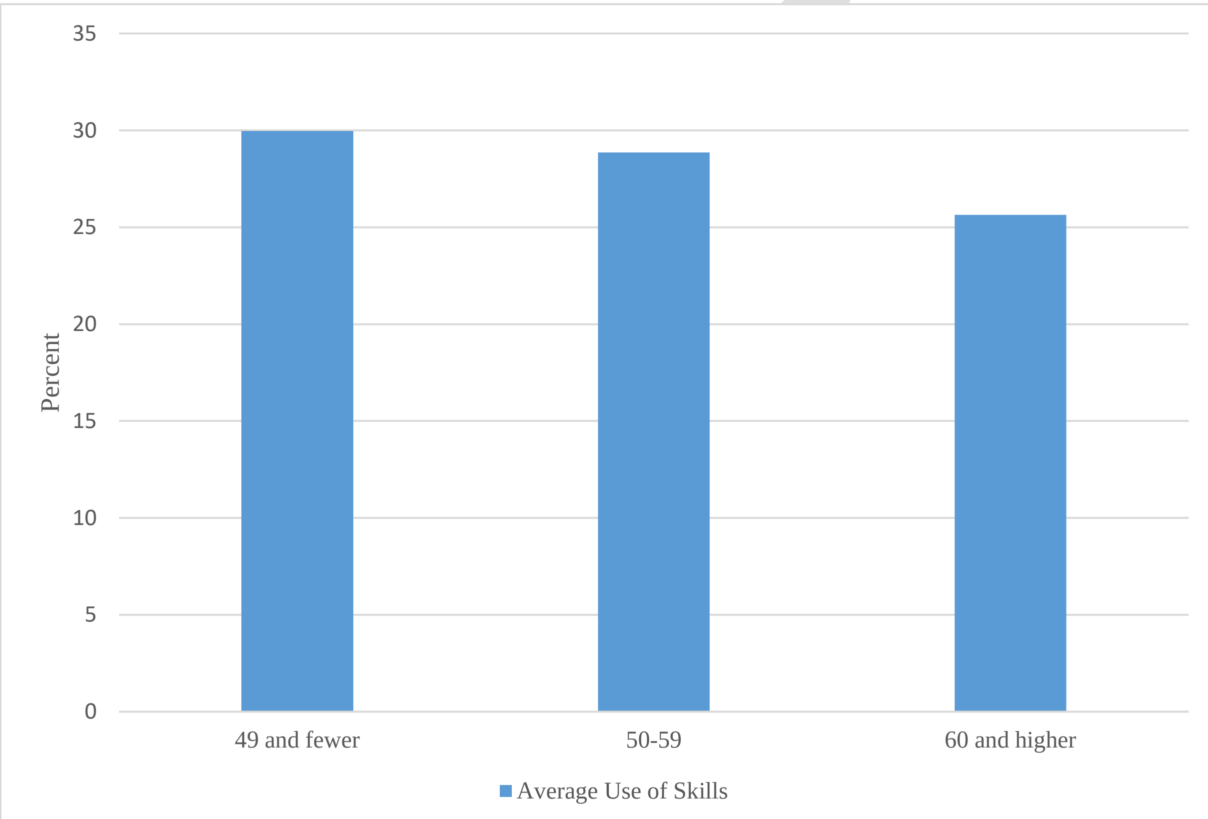
Figure 5. Skill Use by Gender



Caseload

A one-way ANOVA was conducted to determine if the average use of STARR skills was different for groups with different active caseload sizes. Officers were classified into three groups: those with caseloads consisting of 49 or fewer probationers (n = 12), 50 to 59 probationers (n = 16), and 60 or more probationers (n = 13). The average use of skills was highest among officers with a caseload of 49 of fewer probationers (an average of 29.98) and lowest among officers with the largest caseloads of 60 or more probationers (an average of 25.64). However, there were no statistically significant differences in average use of skills between various caseloads $F(2,38) = .608$, $p = .550$.

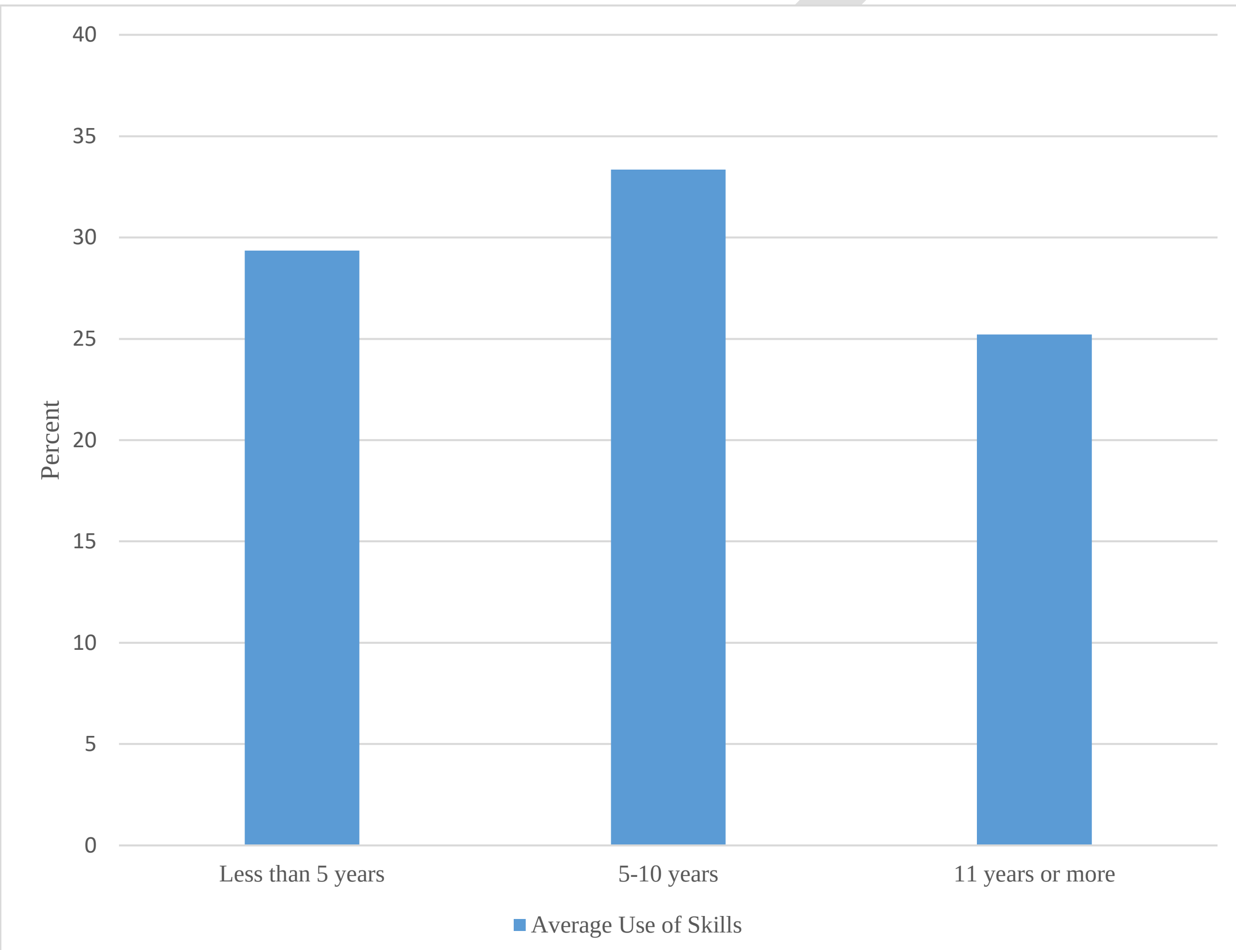
Figure 6. Skill Use by Caseload



Tenure

A one-way ANOVA was conducted to determine if the average use of STARR skills was different based on tenure within MDFL. Officers were classified into three groups: those with a tenure less than 5 years (*n* = 11), 5 to 10 years of experience (*n* = 16), and 11 or more years of experience (*n* = 17). The average use of STARR skills was highest among officers in the middle group (5 to 10 years of experience) (an average of 33.35) and lowest in the group with 11 or more years of experience (an average of 25.22). There were no statistically significant differences in the average use of skills between tenure groups $F(2,41)$, = 2.133, p = .131.

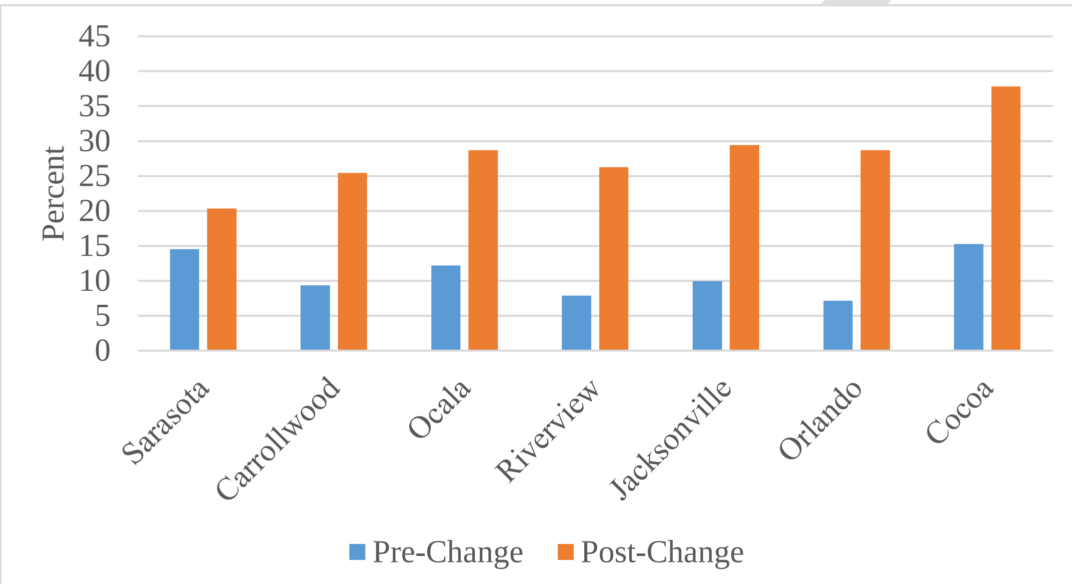
Figure 7. Skill Use by Tenure



How does use of STARR skills vary across offices within the district?

In order to assess whether there was a statistically significant mean difference between the average skill use before and after the policy change within each office, individual paired samples t-tests were run. However, this analysis was not conducted for Fort Myers as the sample size was too small to run bivariate analyses. As illustrated in Figure 8, the proportion of skill use increased in each office examined. This increase in use of skills was statistically significant in all offices except Sarasota. The largest percent increase in use of skills occurred in Cocoa (23% increase) closely followed by Orlando (22% increase).

Figure 8. Skill Use by Office



Sarasota (n = 5)

- Officers used an average of 14.51 STARR skills between September 2018 and February 2019.
- Skill use increased to an average of 20.36 between March 2019 and August 2019.
    - While skill use increased over time, this change was not statistically significant.

Carrollwood (n = 5)

- Officers used an average of 9.34 STARR skills between September 2018 and February 2019.
- Skill use increased to an average of 25.43 between March 2019 and August 2019.
    - The proportion of skill use increased by 16%.

Ocala (n = 5)

- Officers used an average of 12.19 STARR skills between September 2018 and February 2019.
- Skill use increased to an average of 28.70 between March 2019 and August 2019.
    - The proportion of skill use increased by 17%.

Riverview (n = 6)

- Officers used an average of 7.90 STARR skills between September 2018 and February 2019.
- Skill use increased to an average of 26.26 between March 2019 and August 2019.
    - The proportion of skill use increased by 18%.

Jacksonville (n = 9)

- Officers used an average of 9.92 STARR skills between September 2018 and February 2019.
- Skill use increased to an average of 29.42 between March 2019 and August 2019.
    - The proportion of skill use increased by 19%.

Orlando (n = 8)

- Officers used an average of 7.17 STARR skills between September 2018 and February 2019.
- Skill use increased to an average of 28.71 between March 2019 and August 2019.
    - The proportion of skill use increased by 22%.

Cocoa (n = 5)

- Officers used an average of 15.25 STARR skills between September 2018 and February 2019.
- Skill use increased to an average of 37.80 between March 2019 and August 2019.
    - The proportion of skill use increased by 23%.

Table 2. Skill Use by Office

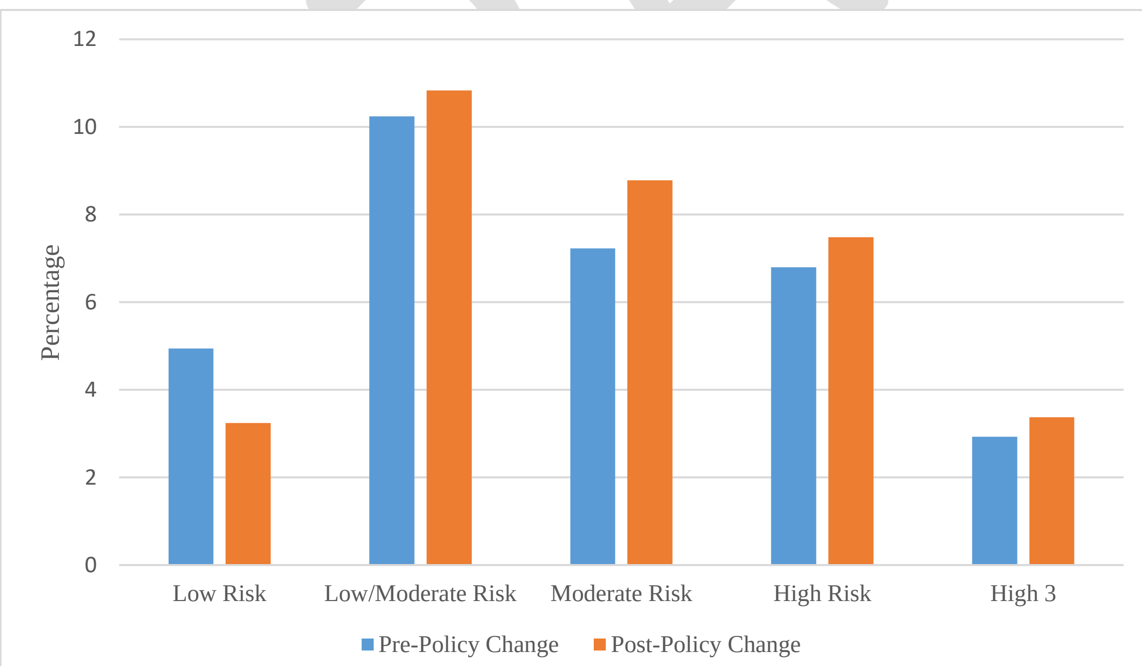
Office (*n*)	**Pre-Change**	**Post-Change**				
	M (*SD*)	*M* (*SD*)	Mean Difference [95% CI]	*Cohen's d*	*t*	*p*
Carrollwood (5)	9.3 (4.4)	25.4 (7.6)	16.1 [8.3, 23.9]	2.6	5.7	*.005*
Cocoa (5)	15.3 (2.2)	37.8 (4.9)	22.6 [13.76, 31.3]	3.2	7.1	*.002*
Jacksonville (9)	9.9 (9.0)	29.4 (4.0)	19.5 [15.0, 24.0]	3.3	9.9	*.000*
Ocala (5)	12.2 (8.2)	28.7 (6.8)	16.5 [9.4, 23.7]	2.9	6.4	*.003*
Orlando (8)	7.2 (11.4)	28.7 (23.3)	21.5 [5.6, 37.5]	1.2	3.2	*.015*
Riverview (6)	7.9 (6.3)	26.3 (7.3)	18.6 [9.4, 27.3]	2.1	5.2	*.003*
Sarasota (5)	14.5 (13.1)	20.4 (13.7)	6.4 [6.1 ,18.8]	0.6	1.4	*.231*

With which risk groups are officers most likely to use STARR skills?

Next, independent t-tests were used to examine the frequency of STARR use with individuals in each risk category before and after the policy change. Risk levels were collapsed into five categories based on MDFL contact standards. The *highest risk* includes those assessed as “high 3”; *high risk* includes “high 2”, “high 1”, and “moderate 3”; *moderate risk* includes “moderate 2”, “moderate 1”, and low/moderate 3”; *low/moderate* includes “low/moderate 2”, “low/moderate 1”, and “low 3”; and *low* includes “low 2” and “low 1”. Data were not available on the total number of contacts each individual officer had within each risk group by month (this was only available for STARR contacts). Thus, the results below demonstrate the total number of STARR contacts for each risk category divided by the total number of personal contacts. While this provides some information regarding who USPOs are using STARR with, it does not capture the total opportunities an officer had to use STARR with individuals in each risk group. As such, the results should be interpreted with caution.

On average, officers were most likely to use STARR skills with individuals assessed as low/moderate risk, followed by moderate risk. Officers were least likely to use STARR skills with individuals assessed as the highest risk. After the policy change requiring STARR users to use a minimum of 8 STARR skills per month, officers increased their use of STARR skills with clients assessed as high, moderate, and low/moderate risk. A statistically significant increase in use of STARR skills was only seen with individuals assessed as moderate risk, with officers using STARR skills with moderate risk clients on average 1.55% more often. After the policy change, STARR users were less likely to use STARR skills with low risk individuals. On average, officers used STARR with low risk individuals 1.7% less often.

Figure 9. Skill Use by Risk Level



Highest Risk (High 3)

- Officers used an average of 2.93 STARR skills with the highest risk individuals between September 2018 and February 2019.
- This proportion increased to an average of 3.37 between March 2019 and August 2019.
    - While average use of STARR skills with the highest risk clients increased slightly, this was not a significant increase.

High Risk

- Officers used an average of 6.79 STARR skills with high risk individuals between September 2018 and February 2019.
- This proportion increased to an average of 7.49 between March 2019 and August 2019.
    - While average use of STARR skills with high risk clients increased slightly, this was not a significant increase.

Moderate Risk

- Officers used an average of 7.23 STARR skills with moderate risk individuals between September 2018 and February 2019.
- This proportion increased to an average of 8.78 between March 2019 and August 2019.
    - Officers significantly increased their average monthly STARR skill use with moderate risk clients by 1.55%

Low/Moderate Risk

- Officers used an average of 10.24 STARR skills with the highest risk individuals between September 2018 and February 2019.
- This proportion increased to an average of 10.83 between March 2019 and August 2019.
    - While average use of STARR skills with low/moderate risk clients increased slightly, this was not a significant increase.

Low

- Officers used an average of 4.94 STARR skills with the highest risk individuals between September 2018 and February 2019.
- This proportion decreased to an average of 3.25 between March 2019 and August 2019.
    - Officers significantly decreased their average monthly STARR skill use with low risk clients by 1.70%

What are the most frequently used STARR skills?

The data were analyzed to assess what skills officers are most likely to use when they implement STARR. Additionally, these analyses help to contextualize changes in skill use to determine if increases in use of STARR after the policy change are due to increases in a specific skill area(s).

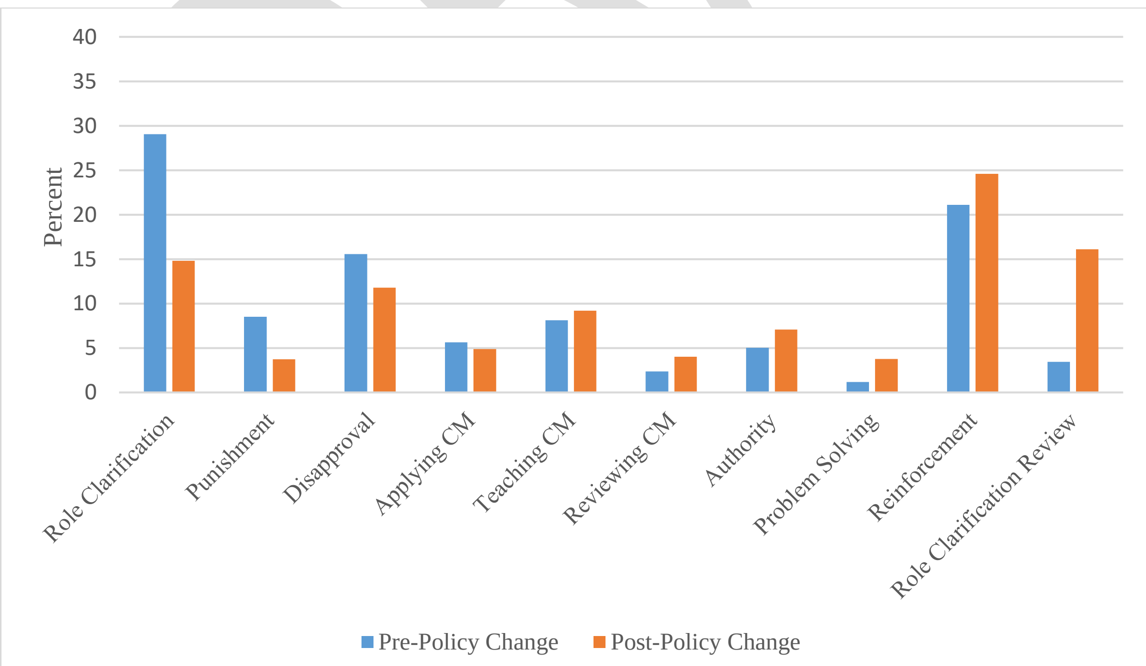
Pre-Policy Change:

- Most frequently used skills:
  - Role Clarification
  - Effective Use of Reinforcement
  - Effective Use of Disapproval
- Least commonly used skills:
  - Problem Solving
  - Reviewing the Cognitive Model
  - Role Clarification Review

Post-policy change:

- Most frequently used skills:
  - Effective Use of Reinforcement
  - Role Clarification Review
  - Role Clarification
- Least frequently used skills:
  - Effective Use of Punishment
  - Problem Solving
  - Reviewing the Cognitive Model

Figure 10. Skill Use by Type



Role Clarification

- Officers used Role Clarification an average of 29.06 times between September 2018 and February 2019.
- This proportion decreased to an average of 14.82 between March 2019 and August 2019.
    - Officers significantly decreased their use of Role Clarification by 14%.

Effective Use of Punishment

- Officers used Effective Use of Punishment an average of 8.51 times between September 2018 and February 2019.
- This proportion decreased to an average of 3.75 between March 2019 and August 2019.
    - Officers significantly decreased their use of Effective Use of Punishment by 5%.

Effective Use of Disapproval

- Officers used Effective Use of Disapproval an average of 15.58 times between September 2018 and February 2019.
- This proportion decreased to an average of 11.80 between March 2019 and August 2019.
    - Officers significantly decreased their use of Effective Use of Disapproval by 4%.

Applying the Cognitive Model

- Officers used Applying the Cognitive Model an average of 5.62 times between September 2018 and February 2019.
- This proportion decreased to an average of 4.90 between March 2019 and August 2019.
    - This decrease in frequency was not a significant change.

Teaching the Cognitive Model

- Officers used Teaching the Cognitive Model an average of 8.13 times between September 2018 and February 2019.
- This proportion increased to an average of 9.20 between March 2019 and August 2019.
    - While average use increased slightly, this was not a significant increase.

Reviewing the Cognitive Model

- Officers used Reviewing the Cognitive Model an average of 2.37 times between September 2018 and February 2019.
- This proportion increased to an average of 4.03 between March 2019 and August 2019.
    - Officers significantly increased their use of Reviewing the Cognitive Model by 2%.

Effective Use of Authority

- Officers used Effective Use of Authority an average of 5.02 times between September 2018 and February 2019.
- This proportion increased to an average of 7.06 between March 2019 and August 2019.
    - While average use increased slightly, this was not a significant increase.

Problem Solving

- Officers used Problem Solving an average of 1.18 times between September 2018 and February 2019.
- This proportion increased to an average of 3.76 between March 2019 and August 2019.
  - Officers significantly increased their use of Problem Solving by 3%.

Effective Use of Reinforcement

- Officers used Effective Use of Reinforcement an average of 21.10 times between September 2018 and February 2019.
- This proportion increased to an average of 24.58 between March 2019 and August 2019.
  - While average use increased slightly, this was not a significant increase.

Role Clarification Review

- Officers used Role Clarification Review an average of 3.44 times between September 2018 and February 2019.
- This proportion increased to an average of 16.12 between March 2019 and August 2019.
  - Officers significantly increased their use of Role Clarification Review by 13%.

Table 3. Skill Use by Type of Skill

	Pre-change	**Post-change**				
Skill Type	*M* (*SD*)	*M* (*SD*)	Mean Difference [95% CI]	*Cohen's d*	*t*	*p*
Authority	5.0 (6.1)	7.1 (6.5)	2.0 [-.4, 4.5]	.26	1.7	*.100*
Disapproval	15.6 (11.9)	11.8 (7.7)	-3.9 [-7.0, -.5]	.36	-2.2	*.024*
Problem Solving	1.2 (3.2)	3.8 (3.6)	2.6 [1.6, 3.6]	.84	5.4	*.000*
Punishment	8.5 (9.1)	3.8 (3.5)	-4.8 [-7.5, -2.1]	.54	-3.6	*.001*
Reinforcement	21.1 (14.0)	24.6 (12.7)	3.5 [-1.9, 8.9]	.20	1.3	*.205*
Role Clarification	29.1 (19.9)	14.8 (6.9)	-14.2 [-20.3, -8.2]	.72	-4.7	*.000*
Role Clarification Review	3.4 (5.4)	16.1 (12.0)	12.7 [8.9, 16.4]	1.10	6.9	*.000*
Applying CM	5.6 (5.6)	4.9 (4.3)	-0.7 [-2.5, 1.1]	.12	-0.8	*.424*
Reviewing CM	2.4 (6.7)	4.0 (4.3)	1.7 [.0, 3.3]	.32	2.1	*.044*
Teaching CM	8.1 (7.5)	9.2 (11.9)	1.1 [-2.7, 4.8]	.09	0.6	*.571*

Note: *n* for all variables is 43.

The largest percentage increase in use of skills between pre- and post-policy change was seen in the use of Role Clarification Review (13% increase) and Role Clarification (14% decrease). More notably however, there was a 5% decrease in Effective Use of Punishment and a 4% decrease in Effective Use of Disapproval, with a 3% increase in Problem Solving and a 2% increase in Reviewing the Cognitive Model. This suggests officers did not simply increase use of punishment-oriented skills post policy change, but rather were engaging in the more difficult and time-consuming STARR skills.

PART II: WHAT IS THE IMPACT OF STARR TRAINING ON PROBATIONER OUTCOMES?

Method

The data used in Part II of this STARR outcome evaluation was combined from administrative probationer records obtained from PACTS with data provided about probation officers by the agency. The sampling frame included all probationers who were supervised on federal post-release probation in the MDFL between March 1, 2019 and August 31, 2019 (N = 4,902). There were 331 cases with one or more elements of missing probationer information. The final sample size for this portion of the study, therefore, includes 4,571 unique probationers.

Independent Variable

The independent variable was probation officer training in STARR. There were 101 probation officers identified who supervised at least one eligible probationer on their caseload during the observation period. [1] Table 4 summarizes the characteristics of these officers. As seen in the table, 54.5% (or 55) probation officers were trained in STARR and 45.5% (or 46) were not trained in STARR. Slightly more than half of this sample were male (53). About a third of these officers were supervisors (33) and almost half had 11 or more years of experience with MDFL (39).[2] The officers were spread across eight division office locations, with the most coming from Carrollwood (23), followed by Orlando (20), Riverview (18), Jacksonville (14), Fort Myers (9), Cocoa and Sarasota (both had 6), and Ocala (5). In this study, probationers supervised by officers who received STARR training represented the treatment group and those supervised by officers who did not receive STARR training served as the comparison group. Assignment to STARR training was not random. Rather, MDFL identified officers to participate in small cohorts of training via requests for volunteers, a process that has slowly incorporated more officers across the district over the last two years.

Table 4. Descriptive Statistics of Probation Officers (N = 101)

Measure		**%**	***n***
STARR-trained		54.5	55
Male		52.5	53
Tenure[a]	*Less than 5 years*	33.3	29
	5 to 10 years	21.8	19
	11 or more years	44.8	39
Probation supervisor[b]		35.5	33
Office location	*Carrollwood*	22.8	23
	Cocoa	5.9	6
	Fort Myers	8.9	9
	Jacksonville	13.9	14
	Ocala	5.0	5
	Orlando	19.8	20
	Riverview	17.8	18
	Sarasota	5.9	6

Note: [a] N = 87. [b] N = 93.

[1] Part II analyses include five officers who were excluded from Part I analyses as they no longer worked for the agency at the time of analyses and access to the officer-level informed needed for Part I was no longer available.
[2] Officer demographic information is missing for individuals who no longer worked with MDFL at the time of data collection as records are destroyed upon separation from the agency.

Dependent Variables

There were four dependent variables analyzed in this study. These dichotomous measures included (1) a positive urinalysis (or drug) test, (2) a violation of one's conditions of probation (i.e., a technical violation), (3) a revocation (or termination) from probation, and (4) being an arrest for a new criminal offense between March 1, 2019 and February 29, 2020 (1 = *yes*, 0 = *no*). Given the non-random nature of the treatment group assignment and the theoretical relevance of offender characteristics on criminal behavior, we employed propensity score modeling (PSM) to isolate the effect of being supervised by a probation officer trained in STARR on these outcomes. The use of PSM in the current context is important because this procedure allowed us to identify and match probationers who were supervised by STARR-trained officers to similar probationers who were monitored by officers not trained in STARR. Once matched, we were able to more accurately measure the impact of STARR training on the four dependent variables.

Control Variables

The probationer variables that were used as covariates for matching included gender (1 = *male*, 0 = *female*), race (1 = *white*, 0 = *non-white*), age (measured in years), marital status (1 = *married*, 0 = *not married*), highest educational level (dummy variables for no GED or high school diploma, GED or high school diploma, and some college or more), employment status (1 = *employed*, 0 = *not employed*), most serious current offense[3] (dummy variables for drugs, firearms, property, violent, white collar, and other), and risk for recidivism (dummy variables for low, low/moderate, moderate, and high risk[4]). Table 5 describes the characteristics of the probationers in this study. Approximately 64.3% (or 2,938) of the probationers were supervised by an officer who was trained in STARR. The remaining 35.7% (or 1,633) of these probationers were supervised by an officer who was not trained in STARR. The individuals in this sample were primarily male (84.8%) and more than half were white (56.6%). The mean age of the probationers was 44.5 years old (SD = 12.0) and approximately one-fifth of the sample was married (22.4%). About a quarter of the probationers had not received their GED or high school diploma (27.0%) and most were gainfully employed (75.7%). Nearly half the sample was on probation for a drug-related offense (46.8%) and about a third were scored as either moderate or high risk for recidivism (28.2%). During the one-year observation period, the most frequent negative outcome observed was a technical violation (17.0%), followed by a revocation of probation (8.6%), a positive urinalysis test (8.5%), and an arrest for a new crime (5.0%).

[3] Offense types were coded using the guidelines for the classification of federal offenses from the Pew Research Center (see Appendix A in Lopez & Light, 2009).

[4] Due to the small percentage of probationers scoring as very high risk for recidivism (≈1%), we combined high- and very high-risk into one category.

Table 5. Descriptive Statistics of Probationers (N = 4,571)

Measure		%	n
STARR trained probation officer		64.3	2,938
Male		84.8	3,876
White		56.6	2,585
Mean age (SD)		44.5 (12.0)	
Married		22.4	1,024
Education	*No GED or diploma*	27.0	1,236
	GED or diploma	44.4	2,028
	Some college or more	28.0	1,280
Employed		75.7	3,460
Most serious offense	*Drugs*	46.8	2,138
	Firearms	8.6	391
	Property	4.7	217
	Violent	4.7	215
	White collar	20.3	930
	Other	14.9	680
Risk category	*Low*	33.8	1,543
	Low/moderate	38.0	1,736
	Moderate	16.8	769
	High	11.4	521
Positive urinalysis test		8.5	390
Technical violation		17.0	777
Revocation of probation		8.6	392
Arrest for a new crime		5.0	229

Analytic Plan

To begin, we compared the probationers supervised by STARR trained and non-STARR trained probation officers. We then employed PSM using the one-to-one nearest neighbor method with a caliper to match the probationers from these two groups.[5] Next, we conducted three sets of pre- and post-match analyses to assess the performance of the matching procedure. In doing so, we first calculated the appropriate t-test or Chi-square statistic for each of the measures and examined the percentage of covariates with statistically significant differences ($p \leq .05$). Equivalent comparison groups should possess fewer than 5% of measures with statistically significant differences (Shadish, Cook, & Campbell, 2002). Second, we calculated the standardized percent bias statistic to assess the degree to which the STARR and non-STARR groups differed on each of the observed constructs (e.g., gender, race, age, risk for recidivism).[6] We focused here on two variants of this measure, including the mean percent bias across all of the measures and percentage of covariates that were greater than or equal to 20%. According to Rosenbaum and Rubin (1985), equivalent groups should not possess any covariates with a percent bias of more than 20%, with lower values representing greater group balance. Third, we calculated the Area Under the Curve (AUC) statistic as a sensitivity check to gauge how well the propensity score predicts placement into the treatment group. The closer an AUC value is to .500, the more it can be said that the propensity score can no longer distinguish between treatment and control cases (Campbell & Labrecque, 2018).

[5] The caliper was determined by multiplying the standard deviation of the propensity score by .25 (see Rosenbaum & Rubin, 1985).
[6] We calculated the standardized percent bias using Austin's (2011) two formulas for continuous and dichotomous measures.

Once matched, we compared the outcomes between the two groups and conducted Chi-square tests to identify if any statistically significant differences exist. We further calculated the Phi (ϕ) statistic to assess the magnitude of the relationship between group placement and the dependent measures. Using Cohen's (1988) guidelines, we interpreted the absolute values of ϕ = .10, .30, and .50 as indicative of small, medium, or large relationships, respectively. Additionally, we evaluated the differences in group outcomes when separated by probationer risk level (i.e., low, low/moderate, moderate, and high risk). Finally, we performed a series of multivariate logistic regression analyses as robustness checks to assess the influence of being supervised by a STARR trained probation officer on the odds of each of the four dichotomous outcomes while controlling for the other theoretically relevant covariates of criminal behavior (e.g., gender, race, age, risk for recidivism).

PART II: RESULTS

Tables 6 and 7 provide a descriptive comparison of the probationers supervised by the STARR and non-STARR trained probation officers before and after the PSM matching procedure. As evident in Table 6, the pre-matched groups differed greatly on most of the covariates examined. Of particular noteworthiness, however, is that the probationers supervised by STARR trained officers were of much greater risk for recidivism compared to those monitored by the untrained officers. Probationers in the STARR trained officer group were three times more likely to be scored as high risk and two and a half times less likely to be scored as low risk than those in the untrained group. In addition, the STARR trained group was also comprised of 10.9% more non-whites, 9.3% fewer people who were married, and 5.5% more individuals without a GED or high school diploma.

Table 6. Pre-Match Group Comparisons and Balancing Statistics

Measure		**% STARR (*N* = 2,938)**	**% Non-STARR (*N* = 1,633)**	***p*-value**	**% Bias**
Male		85.6	83.3	.034	6.4
White		52.7	63.6	<.001	22.2
Mean age (SD)		43.8 (11.8)	45.7 (12.2)	<.001	13.1
Married		19.1	28.4	<.001	22.0
Education	*No GED or diploma*	29.0	23.5	<.001	12.5
	GED or diploma	45.5	42.3	.032	6.5
	Some college or more	24.9	33.6	<.001	19.2
Employed		75.5	76.0	.724	1.2
Most serious offense	*Drugs*	47.3	45.8	.328	3.0
	Firearms	9.9	6.1	<.001	14.0
	Property	4.4	5.5	.096	5.1
	Violent	5.2	3.7	.021	7.3
	White collar	17.3	25.8	<.001	20.8
	Other	15.9	13.1	.012	8.0
Risk category	*Low*	21.7	55.4	<.001	73.8
	Low/moderate	43.6	27.9	<.001	33.2
	Moderate	19.7	11.7	<.001	22.1
	High	14.9	5.0	<.001	33.5

Note: STARR = Strategic Training Aimed at Reducing Recidivism. % Bias = % standardized bias statistic. SD = standard deviation.

These substantive differences make the direct comparison of outcomes between these two groups challenging. Failure to account for these well-known correlates of criminal behavior would inevitably lead to biased results. To address this limitation, we employed a one-to-one PSM matching approach to establish a more equivalent counterfactual comparison group. This research strategy identified 1,368 individuals from each of the treatment and control groups (*N* = 2,736) who were much more comparable on these observed characteristics (see Table 7). This procedure was important because it helps rule out the potential influence that group differences in these covariates may have on the outcomes.

Table 7. Post-Match Group Comparisons and Balancing Statistics

Measure		**% STARR (*N* = 1,368)**	**% No STARR (*N* = 1,368)**	***p*-value**	**% Bias**
Male		82.5	85.4	.042	7.9
White		60.0	60.7	.696	1.4
Mean age (SD)		44.4 (12.3)	45.7 (12.0)	.005	8.7
Married		24.6	26.5	.273	4.4
Education	*No GED or diploma*	23.5	25.7	.169	5.1
	GED or diploma	43.1	43.6	.758	1.0
	Some college or more	32.9	30.1	.118	6.0
Employed		76.2	75.9	.858	0.7
Most serious offense	*Drugs*	42.1	48.8	<.001	13.5
	Firearms	6.8	6.9	.940	0.4
	Property	6.0	4.8	.176	5.3
	Violent	4.8	4.2	.459	2.9
	White collar	23.7	23.2	.752	1.2
	Other	16.7	12.1	.001	13.1
Risk category	*Low*	46.7	46.7	.467	0.0
	Low/moderate	30.6	33.3	.129	5.8
	Moderate	13.6	14.0	.740	1.2
	High	9.1	6.0	.002	11.8

Note: STARR = Strategic Training Aimed at Reducing Recidivism. % Bias = % standardized bias statistic. SD = standard deviation.

Table 8 compares the summary statistics between the treatment and comparison groups to assess for balance or similarities on the range of covariates pre- and post-match. Although the use of PSM successfully reduced the bias between these two groups, there were still some measures that retained a statistically significant difference at the .05 level following the match (e.g., gender, age, drug offense, other offense, and high risk). However, it is important to emphasize that the percentage of measures with a statistically significant difference dropped by 44.4% after the matching procedure was employed. The mean standardized percent bias was also reduced from 18.4 before the match to 5.0 after the match. The application of PSM further eliminated the presence of any measure with a percent bias of 20 or more. Finally, the AUC values predicting group placement were lowered from .697 pre-match to .530 post-match. In totality, these comparative analyses indicate that we achieved a better group balance through the use of PSM. Absent the ability to employ random assignment to the group conditions, these findings provide greater confidence in our ability to isolate the impact of STARR on the probationer outcomes.

Table 8. Average Model Balance Summary, Pre- and Post-Match

Balance measure	Before PSM	After PSM
Percent of statistically significant differences	72.2	27.8
Mean % bias	18.4	5.0
Percent of % bias > 20	38.9	0.0
AUC	.697	.530

Note: PSM = propensity score matching. % bias = % standardized bias statistic. AUC = area under the curve.

Next, we examined the twelve-month post-match outcomes between probationers supervised by officers with and without STARR training (see Table 9). These analyses uncovered a statistically significant difference in the number of technical violations and revocations of probation between these two groups ($p \leq .05$). More specifically, there were 4.7% more technical violations and 2.2% fewer probation revocations found among the STARR trained group relative to the untrained group. Although not statistically significant at the .05 level, there were also 1.6% more positive urinalysis tests and 0.4% fewer arrests found among the probationers in the trained group. The magnitude of these relationships (ϕ = .01 to .07) are all considered small by Cohen's (1988) guidelines.

Table 9. Post-Match Differences in Twelve-Month Outcome Measures

Measure	% Trained (N = 1,368)	% Not Trained (N = 1,368)	$\chi^2(1)$	p	ϕ
Any positive urinalysis test	7.9	6.3	2.69	.101	.03
Any technical violation	17.5	12.8	12.00	.001	.07
Any revocation of probation	7.1	9.3	4.38	.036	.04
Any arrest for new offense	4.4	4.8	0.30	.584	.01

Note: STARR = Strategic Training Aimed at Reducing Recidivism.

Given the imbalance on some of the covariates following the matching procedure, we conducted a series of additional multivariate logistic regression analyses predicting the four outcomes with the host of probationer variables entered as controls (see Table 10). Even when accounting for these covariates, the results remained consistent with those derived from the PSM analyses reported above. The only two models to detect statistically significant relationships between probation officer training group and outcome at the .05 level were those predicting technical violations and probation revocations. Again, these models indicated that probationers supervised by STARR trained officers were more likely to receive a technical violation and less likely to have their probation revoked compared to those supervised by untrained officers. Additionally, the other models revealed evidence of an increase in positive urinalysis tests and a decrease in new arrests for the trained group relative to the untrained group, although neither was found to be statistically significant.

Table 10. Logistic Regression Predicting Probationer Outcomes, Matched Sample (N = 2,736)

Measure		Positive UA	Technical Violation	Probation Revocation	New Arrest
STARR trained probation officer		1.29	1.42^{**}	$.73^{*}$	.87
Male		.76	.76	1.41	.83
White		1.24	1.07	.75	.83
Age		$.98^{**}$	$.98^{***}$	.99	$.95^{***}$
Married		.77	.74	.97	1.26
Education[a]	*GED or diploma*	1.28	1.35^{*}	1.35	1.03
	Some college or more	1.08	1.22	1.16	1.08
Employed		.79	.91	.94	1.01
Most serious offense[b]	*Firearms*	2.43^{***}	1.74^{**}	1.70^{*}	1.16
	Property	1.59	1.73^{*}	1.63	1.10
	Violent	.81	.93	.73	.60
	White collar	1.13	1.61^{**}	1.20	.69
	Other	.68	.97	1.06	.65
Risk category[c]	*Low/moderate*	5.40^{***}	2.69^{***}	1.64^{**}	3.07^{***}
	Moderate	13.52^{***}	4.81^{***}	1.89^{**}	6.63^{***}
	High	17.68^{***}	6.00^{***}	2.31^{**}	6.19^{***}
Constant		$.04^{***}$	$.16^{***}$	$.07^{***}$	$.20^{**}$
Model chi-square (*df*)		235.59 (*16*)	217.39 (*16*)	52.35 (*16*)	110.70 (*16*)
-2 Log likelihood		1,165.12	2,111.58	1,497.95	911.06
Nagelkerke R^2		.206	.133	.044	.127

Note: Reported values are odds ratios. [a] Reference category is no GED or high school diploma. [b] Reference category is drug crimes. [b] Reference category is low risk. $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

Next, we examined the twelve-month post-match outcomes between the STARR trained and untrained officer groups which are separated by probationer recidivism risk level (see Table 11). These moderator analyses suggest that staff training in STARR produced a differential effect among probationers based upon their risk classification. Among the four risk categories examined here, those who were scored as high risk appeared to benefit the most from being supervised by a STARR trained probation officer. High risk probationers in the trained group were 12.1% less likely to have a positive urinalysis test during the twelve-month follow-up period (18.5% compared to 30.6%; $p \leq .05$). Although not statistically significant at the .05 level, probationers supervised by trained officers also had 7.2% fewer technical violations, 5.8% fewer probation revocations, and 6.5% fewer new arrests compared to those monitored by untrained officers. All four of these group differences fell within the small effect size range (ϕ = .08 to .14).

Table 11. Post-Match Differences in Twelve-Month Outcome Measures

Measure	% Trained	% Not Trained	χ^2 (1)	*p*	ϕ
High risk	(*n* = 124)	(*n* = 82)			
Any positive urinalysis test	18.5	30.6	3.94	.047	.14
Any technical violation	30.6	37.8	1.14	.287	.08
Any revocation of probation	11.3	17.1	1.41	.236	.09
Any arrest for new offense	8.1	14.6	2.23	.135	.10
Moderate risk	(*n* = 186)	(*n* = 192)			
Any positive urinalysis test	17.2	17.2	<0.01	.997	.00
Any technical violation	29.0	25.0	0.78	.377	.05
Any revocation of probation	13.4	9.4	1.55	.213	.06
Any arrest for new offense	13.4	7.8	3.16	.075	.09
Low/moderate risk	(*n* = 418)	(*n* = 455)			
Any positive urinalysis test	9.1	5.7	3.66	.056	.07
Any technical violation	17.5	16.7	0.09	.765	.01
Any revocation of probation	7.4	11.6	4.49	.034	.07
Any arrest for new offense	3.6	6.6	4.02	.045	.07
Low risk	(*n* = 639)	(*n* = 639)			
Any positive urinalysis test	2.3	0.3	10.08	.002	.09
Any technical violation	11.7	3.1	34.40	<.001	.16
Any revocation of probation	4.1	6.6	3.98	.046	.06
Any arrest for new offense	1.6	1.4	0.05	.817	.02

Findings among the other risk groups were mixed, with some outcomes that favored the trained group and others that favored the untrained group or produced a negligible difference between the two groups. Among the moderate risk probationers, for example, there was very little difference in the number of positive urinalyses tests between those who were supervised by a STARR trained officer compared to those who were monitored by an untrained officer. The comparisons on the other outcome measures all favored the untrained group. However, all of these differences fell within the small effect size range (ϕ = .00 to .09).

Low/moderate risk probationers supervised by trained officers had 4.2% fewer probation revocations and 3% fewer new arrests compared to those monitored by the untrained officers ($p \leq .05$). Although not statistically significant at the .05 level, the low/moderate risk probationers in the trained group were 3.4% more likely to have a probation revocation and there was little difference in the number of technical violations between groups. All of these differences fell within the small effect size range (ϕ = .04 to .07).

Finally, the impact of officer training in STARR appeared least beneficial to low risk probationers. Only one of the four outcome measures examined displayed an advantage for the trained group. More specifically, probationers supervised by trained officers were 2.5% less likely to have their probation revoked compared to those monitored by untrained officers (p = .046). However, probationers in the trained group were 8.6% more likely to have a technical violation and 2% more likely to have a positive urinalysis test ($p < .01$). The difference in new arrests slightly favored the untrained group, but the difference was substantively trivial.

Owing to the large difference in the number of probationers supervised by STARR trained probation officers (*n* = 2,938) compared to those monitored by untrained officers (*n* = 1,633) in this jurisdiction, we were unable to locate comparable control cases for a significant proportion of treatment cases. This difficulty was further exacerbated by the considerable discrepancies found among the covariates between these two groups (refer to Table 6). Nevertheless, our one-to-one nearest neighbor PSM matching procedure was able to identify 1,368 (or 83.8%) of the probationers in the untrained group who were comparatively similar to 1,368 (or 46.6%) of those in the trained group on the observed covariates (refer to Table 7). Although this matching procedure allowed us to reduce the bias on these factors between the two groups and isolate the impact of officer training in STARR on the four probationer outcomes of interest, this procedure nevertheless excluded 1,570 (or 53.4%) of the probationers monitored by a STARR trained probation officer. Table 12 compares the covariates among the trained group who were successfully matched to those who were excluded in this process. As can be seen in this table, these two groups differed significantly across nearly every dimension examined. Most notably, the non-matched group was more likely than the matched group to be male, non-white, younger, not married, less educated, and higher risk to reoffend.

Table 12. Descriptive Statistics of the Matched and Unmatched STARR Cases

Measure		**% Matched (*n*) (*N* = 1,368)**	**% Non-Matched (*n*) (*N* = 1,570)**
Male		***82.5 (1,129)	88.3 (1,387)
White		***60.0 (821)	46.2 (726)
Mean age (SD)		**44.4 (12.3)	43.2 (11.4)
Married		***24.6 (337)	14.2 (223)
Education	*No GED or diploma*	***23.5 (321)	33.8 (531)
	GED or diploma	*43.1 (589)	47.7 (749)
	Some college or more	***32.9 (450)	17.9 (281)
Employed		76.2 (1,042)	75.0 (1,177)
Most serious offense	*Drugs*	***42.1 (576)	51.8 (814)
	Firearms	***6.8 (93)	12.6 (198)
	Property	***6.0 (82)	2.9 (46)
	Violent	4.8 (65)	5.7 (89)
	White collar	***23.7 (324)	11.8 (185)
	Other	16.7 (228)	15.2 (238)
Risk category	*Low*	***46.7 (639)	0.0 (0)
	Low/moderate	***30.6 (418)	55.0 (863)
	Moderate	***13.6 (186)	25.0 (392)
	High	***9.1 (124)	20.1 (315)

Note: $^{*}p \le .05$. $^{**}p \le .01$. $^{***}p \le .001$.

Given the loss of treatment and control cases through the use of PSM and the number of significant differences found on the covariates between the matched and un-matched trained groups, we next conducted a series of multivariate logistic regression analyses predicting the four probationer outcomes using the full sample of probationers (see Table 13). This strategy allowed us to retain all 4,571 probationers in our analytical models while controlling for the other known predictors of criminal behavior. The results of these analyses were substantively similar to those presented above. To summarize, the probationers in the trained officer group were less likely to have their probation revoked or be arrested for a new crime but were more likely to have a positive urinalysis test and receive a technical violation than those in the untrained group. These findings provide greater confidence in the results of our PSM analyses.

Table 13. Logistic Regression Predicting Probationer Outcomes, Full Sample (*N* = 4,571)

Measure		**Positive UA**	**Technical Violation**	**Probation Revocation**	**New Arrest**
STARR trained probation officer		1.22	1.29**	.78*	.85
Male		.98	.77*	1.44*	.83
White		1.18	1.02	.85	.81
Age		.99**	.98***	.99	.96***
Married		.78	.66***	.90	1.20
Education[a]	*GED or diploma*	1.22	1.16	1.45**	.86
	Some college or more	1.11	1.09	1.18	.70
Employed		.79	.86	.92	.87
Most serious offense[b]	*Firearms*	1.64**	1.53***	1.18	.96
	Property	1.63	1.70**	1.41	.92
	Violent	.54*	1.03	.67	.67
	White collar	1.24	1.83***	1.19	.89
	Other	.58**	1.01	.82	.60*
Risk category[c]	*Low/moderate*	6.67***	2.46***	1.70***	3.21***
	Moderate	15.4***	4.62***	2.32***	6.41***
	High	19.83***	5.65***	3.39***	5.98***
Constant		.02***	.18***	.06***	.17***
Model chi-square (*df*)		339.86 (*16*)	359.90 (*16*)	85.39 (*16*)	159.13 (*16*)
-2 Log likelihood		2,325.72	3,807.59	2,589.67	1,658.34
Nagelkerke R^2		.16	.13	.04	.10

Note: Reported values are odds ratios. [a] Reference category is no GED or high school diploma. [b] Reference category is drug crimes. [b] Reference category is low risk. $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

DISCUSSION

Recommendation #1: Continue to build a culture supportive of STARR and develop strategies to incentivize integration into regular personal contacts.

Data analyses in Part I suggests that the policy implemented in March 1, 2019 to increase use of STARR skills succeeded in increasing USPO use of STARR skills in their personal contacts. Prior to the policy change, officers were using STARR skills in just 5% of all personal contacts (September 2018), compared 26% to 30% of all personal contacts from March 2019 through August 2019. However, in examining these trends, the frequency of use peaked in the month immediately following the policy change and then leveled out to approximately 26% over the following five months. Similar trends were seen in the data exploring the percentage of USPOs who used STARR skills a minimum of eight times per month. The greatest adherence to this policy was seen in March 2019 (86%), however this frequency declined in the following five months and was at 73% as of August 2019.

While not unexpected, this raises an issue to continue to focus on during implementation team meetings. Active work will need to go into continuing to incentivize use and adherence to STARR policies and goals until it becomes standard practice and embedded into the organizational culture. Key targets to support culture change include increasing staff knowledge and understanding for the purpose of STARR and providing ample opportunities to practice skills to increase staff comfort with using the skills until it becomes routine. Continuing to think about formalizing expectations associated with STARR can help in these efforts. Tying performance reviews and raises to use of STARR and making these changes visible can promote implementation and adherence and make it clear to users "what is in it for me" (Pitman, 1994). Additionally, continuing to include front-line users into decisions regarding STARR policy is an extremely valuable tool to support implementation efforts.

Organizational change literature consistently supports the reliance on "change champions" and credits them for helping change efforts to succeed (Beatty & Gordon, 1991; Maidique, 1980; Schon, 1963). These are typically individuals who are skilled in three key areas: initiating, facilitating, and implementing change (Warrick, 2009). The existing implementation team provides an excellent opportunity to bring together "high-impact people" (Warrick, 2009, p. 17) to serve as the change champions of STARR across the district. While the agency is already doing this, it is recommended they continue to assess implementation team members and ensure they are STARR champions. Additionally, given the finding that use of STARR skills was lowest amongst females and more tenured staff, it could be beneficial to identify one champion from each of these groups to inspire and further promote implementation. Lastly, research on organizational change consistently cites the importance of management (both middle and upper levels) demonstrating long-term commitment to change. In doing so, management must play an active role in education, communication, and participation in addition to supporting and incentivizing staff (Pierce et al., 2002). The results presented illustrated that supervisors participated in an average of 30 minutes of booster training, which suggests a potential target for improving the support of the STARR implementation across the district.

With regards to use of STARR by risk level, while the best data were not available to truly understand the proportion of officer contacts integrating STARR by risk, the data presented demonstrates that on average, the most frequent use of STARR skills during the study period occurred with individuals assessed as low/moderate risk, followed by moderate risk. Officers were least likely to use STARR skills with individuals assessed as the highest risk. Research continues to support the use of intensive services and interventions with higher risk individuals (Andrews & Bonta 2010; Andrews & Dowden 2006; Bonta & Andrews, 2007; Gendreau, Goggin & Little 1996; Lipsey & Cullen 2007), with highest risk receiving the highest dosage of intervention, moderate risk receiving a lower dose (Bourgon & Armstrong, 2005; Gendreau & Goggin, 1996; Sperber, Latessa, & Makarios, 2013) and low risk receiving little to no intervention (Andrews & Bonta, 2010; Lowenkamp, Latessa, & Holsinger, 2006). Focusing implementation in this way aids in devoting limited resources (e.g., officer time) to provide services to those who need them most. Overall, data presented in this report suggest that when pushed to implement STARR more frequently, officers followed best practice standards by increasing their use with moderate risk individuals and decreasing their use with low risk offenders. However, the implementation team should consider strategies to encourage officers to increase their use of skills with the highest risk individuals.

Recommendation #2: Consider how individual-level and office-level characteristics impact use of STARR

Findings from Part I highlight some individual-level differences in use of STARR skills. Analyses found that males use STARR skills 10% more frequently in their personal contacts compared to females. Not surprisingly USPOs with smaller caseloads use STARR more frequently compared to those with larger caseloads. Future research should examine this relationship more thoroughly to understand how the make-up of officer caseloads (e.g., by risk level, special populations/needs) may interact with caseload size in influencing officer use of STARR. With regards to tenure, USPOs with 5 to 10 years of experience reported the highest frequency of use compared with those over 11 years of tenure or 4 or fewer years of experience. This suggests an area for the implementation team to consider in terms of how STARR is presented to both the newest USPOs and those who have had the most experience engrained in the previous culture of the agency.

Increases in use of STARR skill use after March 1, 2019 was seen in all offices in the district. This positive finding suggests support across the agency not just for STARR, but for agency policy in general. The largest increase was seen in Cocoa (23%) where the average use of skills was also highest pre-policy change. This suggests Cocoa may serve as an example to other offices in terms of support for and implementation of STARR as well as the identification of potential change champions. The implementation team may want to discuss the culture of the Cocoa office, including supervisor support and accountability practices, to see what might be working well there and what may be translated to other offices in the district. Interestingly, Sarasota had the second highest average use of skills pre-policy change but had the smallest increase in use of skills after March 1, 2019. This suggests an additional avenue for the implementation team to consider.

Recommendation #3: Increase training and skill-building for Problem Solving, and Applying, Reviewing, and Teaching the Cognitive Model.

When examining the STARR use by type of skill, several noteworthy findings emerged. After the policy change, USPOs used Effective Use of Reinforcement more frequently and used Effective Use of Punishment and Effective Use of Disapproval less frequently. This suggests that when USPOs are attempting to integrate STARR into their routine casework, they are focusing on positives and emphasizing highlighting those rather than the negatives. This is a positive finding that falls in line with research evidence that highlights the importance of using positive reinforcement and sanctions at a 4:1

ratio (Wodahl et al., 2011) and the ineffectiveness of punitive strategies focusing on compliance to reduce recidivism (MacKenzie, 2006; Petersilia & Turner, 1993).

Not surprisingly, the lowest engagement occurred in the Problem Solving and Reviewing, Applying, and Teaching the Cognitive Model. These are often believed to be the most difficult skills involved in correctional curriculums like STARR as they require engaging in intervention techniques grounded in cognitive-behavioral therapy, which is a clinical skill not traditionally part of supervision and are much more time intensive techniques. However, research also finds that engagement in cognitive-behavioral techniques is an effective strategy to encourage behavior change amongst justice populations (Wodahl et al., 2011). Previous studies on use of core correctional practices, such as those supported by the STARR skills, finds that expanding coaching sessions at least 18 months after initial training improves not only the frequency skills are used but also the proficiency in use of skills (Labrecque & Smith, 2017). Thus, this finding suggests key targets to address in booster sessions and coaching moving forward to increase USPO comfort with attempting these skills, and ultimately proficiency.

Recommendation #4: Continue the use of STARR during contact sessions with probationers.

This investigation provides mixed support for the effectiveness of probation officer training in STARR on probationer outcomes during a twelve-month follow-up period. Most notably, probationers supervised by officers trained in STARR had fewer probation revocations and new arrests compared to those monitored by untrained officers. This result supports the findings form the prior research community supervision models generally (Bonta et al., 2011; Bonta et al., 2019; Labrecque et al., 2015; Latessa et al., 2013) and on STARR specifically as being effective in reducing probationer recidivism (Lowenkamp et al., 2014; Robinson et al., 2011; Robinson et al., 2012).

This study, however, also found that technical violations and positive urinalysis tests were more prevalent among probationers supervised by officers trained in STARR compared to those supervised by untrained officers. This finding is not surprising and is in line with previous research that examines outcomes when community supervision officers pay closer attention to the behaviors of individuals on their caseloads. For example, when individuals are seen more frequently or experience more in-depth visits with their supervising officer, there are more opportunities for detection of technical violations and/or positive drug screens (e.g., Turner & Petersilia, 1992). Although increases in violations or drug tests are not desirable outcomes, it is evidence that officers may be paying more attention to probationer behavior. Arguably, the most important community supervision outcomes are the successful completion of probation and no new criminal behavior or incarceration. On these metrics, STARR appears a promising practice.

Considering that revocations and rearrests were less likely in the STARR group, this suggests that trained officers may employ a variety of responses to behaviors with individuals on their caseloads either in place of, or prior to making a recommendation of revocation to the court. Additionally, it is possible that while a trained officer may engage in STARR with individuals on their caseloads, those individuals may either require a larger dose of intervention (e.g., more STARR contacts) to impact technical violations and/or positive urinalyses, may require more specific substance use intervention, or they may be resistant to officer implementation of STARR skills. Future research should examine these nuances more in-depth to better identify the mechanisms that support successful implementation of STARR.

Recommendation #5: Prioritize the use of STARR with higher risk offenders.

Although the relationships between officer training and probationer outcomes uncovered in this investigation were generally small in terms of magnitude, it is important to note that the size of these relationships fell in line with the prior scholarship in this area (see the meta-analysis by Chadwick et al., 2015). The truth is that correctional scholars and authorities still do not know the conditions under which community supervision models may produce the best effects. Thus, even when an intervention, such as STARR, is grounded in empirical research and supported by existing studies (Lowenkamp et al., 2014; Robinson et al., 2011), there may be specific factors that moderate effectiveness. For example, it is likely that the way in which officers implement STARR may impact outcomes (e.g., specific skills used, rapport with client, fidelity to the STARR model), or it could be that individuals need to experience a certain dosage of intervention hours to have the greatest impact on behaviors.

It is also possible that STARR is most effective with specific types of probationers and less effective for others. An important finding in this study is that the influence of STARR training was found to be more impactful for higher risk probationers and may produce some negative effects for lower risk cases. This finding is supported by a wealth of previous research on core correctional practices and the RNR model, which finds interventions are more effective when delivered to higher risk individuals and can result in increased reoffending when used with lower risk individuals (e.g., Andrews & Bonta, 2010; Andrews, Zinger, et al., 1990; Andrews & Dowden, 1999; Bonta, Wallace Capretta, & Rooney, 2000; Dowden & Andrews, 1999a, 1999b, 2000; Hanley, 2006; Lowenkamp & Latessa, 2002).

The current study found that high risk probationers supervised by STARR trained officer had a 12.1% significant reduction in positive urinalysis, and non-significant declines in technical violations (7.2%), revocations (5.8%), and rearrests (6.5%). And further in line with existing research, STARR was least beneficial for the lowest risk individuals in the current sample. This provides preliminary evidence to support prioritizing resources on higher risk individuals under supervision. However, future research should examine the nuances associated with implementing STARR with various risk groups. For example, perhaps specific STARR skills should be reserved for higher risk groups (e.g., the more challenging problem solving and cognitive model), while more basic STARR skills such as role clarification would be an appropriate level of intervention for lower risk probationers.

Recommendation #6: Continue to develop more robust ways to evaluate the effectiveness of STARR.

Although this evaluation helps advance knowledge about the impact of STARR on probationer outcomes, readers should exercise caution in interpreting and generalizing these results. There are several limitations to this study that should be understood and addressed in future research evaluations. First, this study makes a critical assumption that officers trained in STARR apply these skills with fidelity in their interactions with the probationers they supervise and also that they all do so equally well. This may or not be the case in reality. As evidenced by findings in Part I of this report, the use of STARR skills by trained officers varied across offices and officers. This suggests a need to expand this research to capture more in-depth data about how and under what circumstances STARR is implemented in practice. Trained officers may use skills in all, some, or none of their interactions and variation in their skill level most certainly exists. Future research should seek to unpack how often officers use STARR skills, how proficient they are in those skills, under what circumstances they use specific skills and what impact such usage has on probationer outcomes.

Second, this investigation involved a non-random sample of probationers from the caseloads of federal probation officers spread across the MDFL. We applied PSM to account for the differences in the observed offender covariates, but due to the restraints in the data the matches were not perfect and further excluded a significant number of treatment cases. We applied multivariate logistic regressions as robustness checks on our results and the findings across both strategies were substantively similar. Although the use of PSM and regression provides convergent validity of our study findings, future research should explore additional ways to gather more cases and additional types of covariates for matching. For example, future research may include a quasi-experimental design where data from MDFL are matched to data from another, comparable district to isolate the effects of STARR on probationer outcomes.

Finally, although our study uncovered promising results, it nevertheless involved a relatively short follow-up period of one year. We expect that continued emphasis on STARR training and further implementation of the model across the district will result in more positive effects. To test this prediction, future scholarship should continue to evaluate the effects of officer training on probationer outcomes of a longer duration (e.g., 2 or 3 years). Additionally, it is critical that future research account for fidelity to the STARR model. It is very possible that small effect sizes were detected in the current study due to the inability to account for how well officers implemented STARR. Data on key variables such as quality of STARR skills used and whether the STARR skill used was appropriate for the situation could strongly improve the current evaluation. However, given the current study was conducted early in MDFL's implementation efforts finding even small effects is positive. The next step in this process should be creating measures of quality and fidelity to truly understand the impact of STARR on probationer outcomes. Such effort would not only be a mechanism for detecting effects but would also serve as a means to inform officer performance reviews and incentives.

References

Alarid, L. F., & Jones, M. (2018). Perceptions of offender satisfaction on probation and supervised release with STARR skill sets. *Federal Probation, 82*(1), 37-44.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th Ed.). New Providence, NJ: LexisNexis Matthew Bender.

Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology, 50*(1), 88-100.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.

Battalino, J., Beutler, L., & Shani, A.B. (1996). Large-system change initiative: Transformation I progress at the California Department of Corrections. *Public Productivity & Management Review,* 20(1), 24-44.

Beatty, C., & Gordon, J. R. M. (1991). Preaching the gospel: The evangelists of new technology. *California Management Review, 33*(1), 73-94.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation* (User report 2007-06). Ottawa, Ontario: Public Safety Canada.

Bonta, J., Bourgon, G., Rugge, T., Scott, T. L., Yessine, A. K., Gutierrez, L. K. & Public Safety Canada. (2010). The strategic training initiative in community supervision: Risk-need-responsivity in the real world 2010-01. *Ottawa: Public Safety Canada*.

Bonta, J., Bourgon, G., Rugge, T., Scott, T. L., Yessine, A. K., Gutierrez, L., & Li, J. (2011). An experimental demonstration of training probation officers in evidence-based community supervision. *Criminal Justice and Behavior*, 38(11), 1127-1148.

Bonta, J., Rugge, T., Bourgon, G., & Wanamaker, K. A. (2019). A conceptual replication of the Strategic Training Initiative in Community Supervision (STICS). *Journal of Experimental Criminology, 15*(3), 397-419.

Bonta, J., Wallace-Capretta, S., & Rooney, J. (2000). Can electronic monitoring make a difference? An evaluation of three Canadian programs. *Crime & Delinquency*, *46*(1), 61-75.

Bourgon, G., & Armstrong, B. (2005). Transferring the principles of effective treatment intro a “real world” prison setting. *Criminal Justice and Behavior, 32*(1), 3-25.

Campbell, C. M., & Labrecque, R. M. (2018). *Panacea or poison: Can propensity score modeling (PSM) methods replicate the results from randomized control trials (RCTs)?* Washington, DC: National Institute of Justice.

Chadwick, N., Dewolf, A., & Serin, R. (2015). Effectively training community supervision officers: A meta-analytic review of the impact on offender outcome. *Criminal Justice and Behavior*, 42(10), 977-989.

Clark, P. B., & Wilson, J. Q. (1961). Incentive systems: A theory of organization. *Administrative Science Quarterly, 6*(2), 129-166.

Cohen, J. (1988). *Statistical power and analysis for the behavioral sciences* (2^{nd} ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cullen, F.T. and Gendreau, P. (2000). 'Assessing correctional rehabilitation: Policy, practice and prospects, In J. Horney (ed*), Criminal Justice: Policies, Processes, and Decisions of the Criminal Justice System, Vol. 3*. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Farrell, J. L., Young, D. W., & Taxman, F. S. (2011). Effects of organizational factors on use of juvenile supervision practices. *Criminal Justice and Behavior*, 38(6), 565 -583.

Gendreau, P., & Goggin, C. (1996). Principles of effective correctional programming. *Forum on Corrections, 8*, 38-41.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works!, *Criminology*, *34*(4), 575-608.

Labrecque, R. M., & Smith, P. (2017). Does training and coaching matter? An 18-month evaluation of a community supervision model. *Victims & Offenders, 12*(2), 233-252.

Labrecque, R. M., Smith, P., & Luther, J. D. (2015). A quasi-experimental evaluation of a model of community supervision. *Federal Probation*, 79, 14-19.

Latessa, E. J., Smith, P., Schweitzer, M., & Labrecque, R. M. (2013). *Evaluation of the effective practices in community supervision model (EPICS) in Ohio*. Unpublished manuscript. Center for Criminal Justice Research, University of Cincinnati, OH.

Lin, A. C. (2000). *Reform in the making: The implementation of social policy in prison*. New Jersey: Princeton University Press.

Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science,* 3, 297-320.

Lofquist, E. A., Greve, A., & Olsson, U. H. (2011). Modeling attitudes and perceptions as predictors for changing safety margins during organizational change. *Safety Science, 49*, 531-541.

Lopez, M. H., & Light, M. T. (2009). *A rising share: Hispanics and federal crime.* Washington, DC: Pew Research Center.

Lowenkamp, C. T., Holsinger, A., Robinson, C. R., & Alexander, M. (2014). Diminishing or durable treatment effects of STARR? A research note on 24-month re-arrest rates. *Journal of Crime and Justice, 37*(2), 275-283.

Lowenkamp, C. T., & Latessa, E. J. (2005). Increasing the effectiveness of correctional programming through the risk principle: Identifying offenders for residential placement. *Criminology & Public Policy*, *4*(2), 263-290.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The risk principle in actin: What have we learned from 13,676 offenders and 97 correctional programs? *Crime & Delinquency*, *52*(1), 77-93.

MacKenzie, D. L. (2006). *What works in corrections? Reducing the criminal activities of offenders and delinquents.* Cambridge, U.K.: Cambridge Press.

Maidique, M. A. (1980). Entrepreneurs, champions, and technological innovation. *Sloan Management Review, 21*(2), 59-76.

Oreg, S., & Berson, Y. (2011). Leadership and employees' reactions to change: The role of leaders' personal attributes and transformational leadership style. *Personnel Psychology, 64*, 627-659.

Petersilia, J., & Turner, S. (1993). Intensive probation and parole. *Crime and Justice, 17*, 281-335.

Pierce J. L., Gardner, D., & Dunham, R. B. (2002). *Management and organizational behavior: An integrated perspective.* New York: Thomson Learning.

Pitman, B. (1994). Critical success factors to organizational change. *Journal of Systems Management, 45*(9), 40.

Robinson, C. R., VanBenschoten, S., Alexander, M., & Lowenkamp, C. T. (2011). A random (almost) study of Staff Training Aimed at Reducing Re-Arrest (STARR): Reducing recidivism through intentional design. *Federal Probation*, 75, 57.

Robinson, C. R., Lowenkamp, C. T., Holsinger, A. M., VanBenschoten, S., Alexander, M., & Oleson, J. C. (2012). A random study of Staff Training Aimed at Reducing Re-arrest (STARR): Using core correctional practices in probation interactions. *Journal of Crime and Justice*, *35*(2), 167-188.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33-38.

Rubin, R. S., Dierdorff, E. C., Bommer, W. H., & Baldwin, T. T. (2009). Do leaders reapw hat they sow? Leader and employee outcomes of leader organizational cynicism about change. *Leadership Quarterly, 20*, 680-688.

Rudes, D.S. (2012). Getting technical: Parole officers' continued use of technical violations under California's parole reform agenda. *Journal of Crime & Justice*, 35, 249-268.

Schein, E. (2010). *Organizational culture and leadership.* San Francisco: Wiley.

Schon, D. A. (1963). Champions for radical new inventions. *Harvard Business Review, 41*(2), 77-86.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Cengage: Houghton Mifflin.

Smith, P., Gendreau, P., & Swartz, K. (2009). Validating the principles of effective intervention: A systematic review of the contributions of meta-analysis in the field of corrections. *Victims and Offenders, 4*(2), 148-169.

Sperber, K. G., Latessa, E. J., & Makarios, M. D., (2013). Establishing a risk-dosage research agenda: Implications for policy and practice. *Justice Research and Policy, 15*(1), 123-141.

Taxman, F.S., Henderson, C., Young, D.W., & Farrell, J. (2012). The impact of training interventions on organizational readiness to support innovations in juvenile justice offices. *Administration of Mental Health Policy and Mental Health Services Research, 41*(2), 1-12.

Viglione, J. (2017). Street-level decision making: Acceptability, feasibility, and use of evidence-based practices in adult probation. *Criminal Justice and Behavior, 44*(10), 1356-1381.

Viglione, J., Rudes, D.S., & Taxman, F.S. (2015). Misalignment in supervision: Implementing risk/needs assessment instruments in probation. *Criminal Justice and Behavior, 42*(3), 263-285.

Warrick, D. D. (2009). Developing organization change champions. *Organizational Development Practitioner, 41*(1), 14-19.

Wodahl, E. J., Garland, B., Culhane, S. E., & McCarty, W. P. (2011). Utilizing behavioral interventions to improve supervision outcomes in community-based corrections. *Criminal Justice and Behavior, 38*(4), 386-405.

Young, D.W, Farrell, J, & Taxman, F.S., (2012). Impacts of juvenile probation training models on youth recidivism. *Justice Quarterly, 30*(6), 1-22.