

**A Process and Outcome Evaluation of the  
4-H Mentoring/ Youth and Families with Promise (YFP) Program**

**SUMMARY**

**March 2012**

**Mary E. Poulin and Stan Orchowsky**



**Justice Research and Statistics Association  
777 North Capitol St., NE  
Suite 801  
Washington, DC 20002**

This project was supported by Award No. 2005-IJ-CX-0046 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect the views of the U.S. Department of Justice.

## **INTRODUCTION**

In December 2004, the National Institute of Justice (NIJ) issued a solicitation for evaluation of four youth programs funded by the Office of Juvenile Justice and Delinquency Preventions (OJJDP). One of those programs was Utah's 4-H Mentoring/Youth and Families with Promise (YFP) program. YFP is a statewide 4-H mentoring program designed for at-risk youths who are approximately 10-14 years old. Youths are paired with mentors, participate in 4-H activities, and also attend Family Night Out (FNO) activities to strengthen family bonds. JRSA received the award from NIJ to conduct a process and outcome evaluation of the YFP program.

## **PROCESS EVALUATION**

The purpose of the *process evaluation* was to assess the degree to which the YFP program was implemented and operated as planned. The evaluation covers approximately three years of program operation, from September 2005 through September 2008. During that time, the number and location of program sites, sources of funding, and how program implementation data were collected changed, but the program design remained consistent. Multiple data sources, including official program records and documentation describing the program, reports from program participants and mentors, and observations of program activities were used to conduct the process evaluation.

We began by comparing program guidelines found in the YFP Program Guide with what was actually taking place in programs and found both consistencies and inconsistencies. Home visits were provided as planned; the types of interventions were appropriate; on average sites offered FNO and 4-H activities as required; and youths were matched with mentors as specified in the program guide. Some interventions and requirements did not appear to be implemented as

planned: some sites changed the schedule of program activities during the summer; some youths admitted were younger than the targeted age range; many youths attended FNOs less often than required, but many attended 4-H more often than required; on average youths met with their mentors less frequently per month than expected, though there was substantial variation by site; some mentors did not receive an orientation training or the mentor manual; and early contacts between mentors and site coordinators were less frequent than required for many sites. Finally, though not a requirement or a guideline, few youths were matched with a grand mentor (i.e., older adult); most youths were matched with a young adult mentor.

Availability of program implementation data was a concern. Many sites were doing a good job of reporting such data, but others were reporting little, if any. Several other important variables were not routinely collected by the program, including discharge date; reasons for, and sources of, referral to the program; and information about why a youth was matched with a mentor. We were not clear whether the activities reported reflected reality or record-keeping limitations. It is possible, for example, that mentors met with youths but did not record the meetings in the online log. It is crucial that the program accurately capture all program activities so it can determine where minimal levels of service are not being met.

We examined program logs of reported activities, which showed that youths received an average of 2.35 in-person intervention contacts (meeting with mentor, participating in 4-H, and attending FNO) per month, or just over half the monthly requirement. Though participation in 4-H activities was generally met or exceeded requirements, attendance at FNO activities and meetings between youths and mentors did not. One likely reason for youths' attending more 4-H than FNO activities is that many sites offered many more 4-H activities than recommended by the Program Guide, but, on average, they offered few more than the required number of FNO

activities. Further, for many sites YFP operated primarily as a school-year program; there was less mentoring during the summer months and some sites did not offer 4-H and FNO during the summer. The school-year program schedule also likely accounts for why in some sites the average length of the first match between a mentor and mentee was about 9.5 months. Given that YFP is part of 4-H, the program could continue activities through FNOs and 4-H activities during the summer even if the mentoring relationship were suspended.

There is no presumed length of stay for the program. Youths typically stay in the program as long as they like and few youths graduated from the program during our study. This is a concern if the program is not able to provide services to youths who would benefit because it is serving longtime attendees who no longer require program interventions.

To help sites identify and resolve problems related to youth participation in the program, promote better accountability among mentors, and help sites with program implementation we recommend that they make the following alterations in their data collection practices: 1) site coordinators should monitor mentor contact data submitted in the online log on a monthly basis to ensure data are entered; 2) the state 4-H office should monitor all data submitted by sites monthly to ensure data are entered and accurate; 3) reasons for referral and source of referral should be added to the online log; and 4) fields should be added in the online log to track scheduled versus completed mentor meetings.

## **OUTCOME EVALUATION**

We used a nonequivalent groups design for the outcome evaluation (pre- and post-testing for an experimental group as well as a comparison group) that includes up to a three-year follow-up for youths attending the program. We collected pre- and post-test data from youths entering YFP from 2005 to 2008 as well as their parents (YFP group), and from two cohorts of students

from various schools in grades 4-8 (comparison group). Major components of pre- and post-tests included the Behavioral and Emotional Rating Scale Version 2 (BERS-2), as well as questions measuring self-reported delinquent behaviors and what youths would like to work on (or did work on) with their mentors during the program. The BERS-2 is a 52-item scale designed to assess the behavioral and emotional strengths of children in five domains (subscales): Interpersonal Strength, Intrapersonal Strength, Affective Strength, Family Involvement, and School Functioning. The instrument also contains a supplemental Career Strength scale consisting of five items that measure interest and aptitude for career and vocational development. For youths in the YFP group, the post-test also included questions on mentor relationship quality and perceptions of the YFP program.

To assess the long-term outcomes of program participation and sustainability of program effects, we scheduled follow-up telephone interviews with YFP and comparison group youths to occur annually for up to three years after program discharge/date of the last post-test. Given the program objective associated with improving school performance, we obtained official school records of attendance and grades prior to, during, and up to three years following involvement in the program (or last post-test for the comparison group). Since the program targets at-risk youths, we collected data on YFP and comparison youths' involvement with the juvenile court system prior to, during, and up to a maximum of three years following their involvement in the program/study.

Because youths tended to stay in the program a long time, we collected additional post-test data each year to capture information on any changes occurring. Post-tests were scheduled for approximately 8 (post-test 1), 20 (post-test 2), and 32 (post-test 3) months following the pre-

test. The table below shows the number of completed tests we received for the YFP and comparison group youths.

*Youth Pre- and Post-Test Form Completion by Group*

<b>Number of Forms Received</b>	<b>YFP Group (included in outcome analyses, if different<sup>a</sup>)</b>	<b>Comparison Group</b>
Pre-test	392 (385)	327
Post-Test 1 Only (no Pre-Test)	9	Not applicable
Pre-Test and Post-Test 1	270 (257)	280
Pre-Test and Post-Test 2	98 (96)	144
Pre-Test, Post-Test 1 and Post-Test 2	93	131
Post-Test 3	28	Not applicable

<sup>a</sup>Some forms were not used in the outcome analyses because they were completed too late for inclusion.

The annual follow-up telephone interviews with YFP and comparison group youths mentioned above were used to assess current academic performance, interpersonal competence, family relationships, and delinquency/risky behavior as well as perceptions of YFP program involvement (YFP youths only). The table below shows the number and percentage of follow-up interviews conducted.

*Number of Annual Follow-up Interviews Completed by Group*

	<b>YFP Group (% Eligible Completed)</b>	<b>Comparison Group (% Eligible Completed)</b>
Year One	132 (40%)	103 (37%)
Year Two	107 (31%)	147 (49%)
Year Three	38 (30%)	51 (36%)

We examined the YFP and comparison groups for comparability, and then used the youth version of the BERS-2 subscales to look at changes in the youths from the pre-test to post-tests at approximately 8 and 20 months. There were significant differences between the two groups on a number of socio-demographic and risk factors at the pre-test, with YFP youth more likely to be non-white, younger, living with only one parent, and having one or more self-reported school suspensions in the six months prior. For the BERS-2 subscales, YFP youths had lower average scores and showed greater variability in their scores than youths in the comparison group. Given that the program targets youths who are at-risk and we were not able to randomly assign youths to the YFP and comparison groups, these differences are not unexpected.

Using analyses of covariance (ANCOVA) to control for pre-test differences, analyses of the BERS-2 subscale results showed a significant, positive change on post-test 1 for the YFP group on the *affective strength* measure. That is, there was evidence that regardless of their level of affective strength when they started the program, YFP youths had a greater ability to accept affection and express feelings after being in the program for approximately eight months when compared with comparison youths. No other significant between-group differences were found at post-test 1 for the YFP and comparison groups as a whole. We conducted separate analyses for those YFP youths who received at least two-thirds of the services (mentoring, FNO, and 4-H) they should have received according to program requirements given their time in the program (n

= 76). When these YFP youths were examined, the previously observed difference on affective strength was larger, but failed to achieve statistical significance due to the smaller sample size.

At post-test 2, the only significant difference between the two groups on the BERS-2 was on *family involvement*. Youths in the comparison group had significantly higher family involvement than the youths in the YFP group.

Another set of ANCOVAs was carried out to assess selected youth differences (age, race, gender, and parental living status) that might moderate program effects. The age by group (YFP vs. comparison) interaction was statistically significant for three BERS-2 subscales – School Functioning, Affective Strength, and Career Strength – and the interaction for a fourth subscale, Family Involvement, approached significance. In all cases, younger YFP participants scored higher on the first posttest than older participants, a difference not observed in the comparison group.

Separate analyses using gain scores rather than ANCOVAs showed that mean scores improved significantly for YFP youths for all BERS-2 scales, except Career Strength, at approximately 8 and 20 months (post-test 1 and 2, respectively) of participation. Youths with longer average lengths of stay in the program (29.7 months) had greater improvements on all the BERS-2 scales within approximately eight months of program involvement than youths with a shorter average lengths of stay (14.2 months), and most of the improvement in scores that occurred over 20 months occurred in the first eight months of the program.

The Youth-Mentor Relationship Questionnaire (YMR) was administered at each post-test for youths in the YFP group only. On average, youths rated their relationships as more rather than less successful, but the greatest success was achieved in the absences of negative emotions about the relationship (Not Unhappy subscale), while the least success was achieved with how

the mentors helped youth cope with problems (Helped to Cope subscale). Girls were more likely to report that mentors helped them with their problems, while non-whites were more likely to report negative feelings about their mentors. Neither of two measures of program involvement, length of time or dosage, was correlated with scores on the YRM subscales.

When youths were asked at post-tests 1 and 2 about how much they thought the program helped them, an overwhelming majority (90% and 94%, respectively) indicated that the program helped them “somewhat” or “very much.” At both post-tests 1 and 2, youths reported that mentor activities were the part of the program that helped them the most (47% and 53%, respectively).

With regard to school data, at least some data were available for 39% of the YFP group and 73% of the comparison group. Lack of cooperation from the schools and school districts contacted was the primary reason for missing data. Issues with the quality of these data made precise assessments of differences between YFP and comparison youth difficult. Given these difficulties, we limited our analysis to assessing whether the number of days absent increased, decreased, or remained the same for youths with multiple absences in the period preceding program/study involvement to the last period of study involvement. Comparable proportions of youths in the YFP and comparison groups had decreased absences during this time, though YFP youths were more likely to have an increase in absences.

With regard to court involvement, we collected data prior to, during, and up to a maximum of three years following youths’ involvement in the program/study. Data on court involvement were available for 98% of the comparison group and 82% of the YFP group. Only 3% of youths had a history of court involvement at any time before the study began (2.8% for the comparison group and 3.1% for the YFP group). During the time YFP youths were in the program, a significantly greater proportion (6.2%) had charges brought to court than comparison

group youth (2.2%;  $\chi^2 = 6.66$ ,  $p = .01$ ). YFP youths were more likely than comparison youths to have court involvement for a new offense in both the first and second year following discharge.

We began the examination of follow-up data by comparing the follow-up samples of YFP and comparison group youths with the original samples. As expected, the subsamples that comprised the follow-up groups showed fewer “risk factors” than the overall study samples. Therefore, the results presented should be interpreted with caution, since they may reflect a positive sample selection bias.

When asked at the one-year follow-up how much the 4-H Mentoring/YFP program helped them, 91% of YFP youths indicated that it helped them somewhat or very much. Activities with the mentor (60%) and FNOs (21%) were activities that YFP youths reported helped them the most. When specifically asked about any changes the youths noticed in their family relationships in the year following involvement in YFP, 74% reported some sort of improvement with their family. Though a little less frequently than at the year-one follow-up, most youths reported that the program helped them “somewhat” or “very much” (84%). Activities with the mentor (55%) remained the most commonly selected response when youths were asked what program activity helped them the most.

A series of additional questions asked at follow-up, some of which were taken from the BERS-2 scales, suggest that YFP youths have more self-reported school, family, and social issues in the first year after they left the program than comparison group youths. As noted previously, YFP youth were also more likely to come into contact with the juvenile court during the follow-up period than were comparison group youth.

We collected cost data on the YFP program and found substantial variation by site in terms of the cost per day to serve a youth. We calculated a cost per day of \$6.70 based on the

total expenditures and total number of youths served by all the sites. Given this cost per day in FY 2009, if a youth stayed for 12 months, it cost \$1,825 to serve the youth. We concluded that the 4-H/YFP mentoring program costs are in range with those of similar programs.

## **DISCUSSION AND CONCLUSIONS**

We did not see any evidence that youths in the YFP group had substantially higher school grades or fewer school absences during the program or after two years of having left the program when compared to the comparison group. Since neither official records of school performance nor scores on the School Functioning scale showed significant differences between the treatment and comparison group, we conclude that there is no evidence to suggest that the program improved academic performance.

There is no evidence that the program strengthened family relationships, as indicated by the lack of significant difference in Family Involvement scores between the treatment and comparison groups. There is also no evidence that the 4-H Mentoring/YFP program had an effect on delinquency. Though the proportion of youths who offended was small, YFP youths were more likely to offend while in the program, and one and two years following the program, than comparison youths.

With regard to long-term effects, we found that YFP youth continued to view the program in positive terms after leaving the program. These youth reported that they had benefitted from the program and their relationships with their mentors. However, YFP youth were more likely to have legal charges brought against them up to two years after program involvement, and self-reported more negative school, family and social issues after program completion than comparison youth.

While the summary of findings presented above is certainly not positive, our study did show many positive findings regarding the program. The program appears to have successfully targeted higher risk and/or need youth, as indicated by their lower initial scores when compared with those of the comparison group youth on all components of the BERS-2. Mentor and mentee meetings appeared to take place on a regular basis, and many 4-H activities were offered in most sites. Over half of YFP youth felt that the program had helped them “very much” and the vast majority felt the program had helped them at least somewhat. Almost half of YFP youth named “activities with my mentor” as being the part of the program that helped them the most. Moreover, these positive feelings were reported several years after youth were no longer involved in the program.

Despite the positive findings noted above, we are forced to conclude that the YFP program in general did not produce the desired outcomes. There are several possible explanations (other than methodological ones) for why we failed to observe more positive outcomes. One of the most likely explanations is suggested by the findings of the process evaluation: that not enough youth received enough program services.

Several methodological issues suggest caution when interpreting our findings. Despite our best efforts, we were unable to obtain a comparison group that was equivalent to the YFP group on a number of key characteristics. Generally speaking, the YFP group was in fact riskier and needier than the comparison group youth. This issue is particularly important when one considers that the YFP group’s subscale scores did indeed improve from pre-test to the first post-test. It is possible that if the two groups had started out equivalent, we would have found that the YFP group would have looked better than the comparison group on the outcomes measured.

For a variety of reasons, we were not able to include as many youth as we would have liked in the study. The number of youth served by the YFP program during the time we conducted our data collection was smaller than in previous years, a result of cuts in program funding and some sites closing down. As it became obvious that the sample size was not going to be as large as we had hoped, we used a variety of approaches to increase the numbers, including lengthening the planned data collection period and adding the prior year's (2005) YFP cohort to our own data collection (which we had not planned to do). Nevertheless, it is possible that our sample sizes may simply have been too small for us to detect significant differences between the YFP and comparison groups. In addition, the attrition rates for the follow-up interviews were higher than we had anticipated, and this affected our sample size.

In conducting the process evaluation component of the assessment we encountered a number of issues with the "dosage" data being maintained by the program. The quality of these data may have directly impacted our conclusions from this part of the study, since we based our assessment of service delivery on these records. As noted previously, it is possible that more YFP youth attended more mentor meetings, 4-H activities, and FNOs than were recorded in the log data we received from the site coordinators. If this was the case, our hypothesis that the lack of positive outcomes may have been due to the low program dosage received by YFP youth would be contradicted.

Given these methodological issues and the concerns associated with program implementation that we identified in the process evaluation, it would be unwise for us to make definitive statements about the program's success or failure. At most we can say there is limited evidence of program success, but the program should consider further evaluation once issues with program implementation have been addressed.