CD-27404

# YOUNG MEN AS FATHERS PROGRAM: FINAL REPORT

**Jeff Lara**
**Michael Ezell**

**State of California**
**Department of the Youth Authority**
**Research Division**
**December 1997**

**Prepared under grant number 90CA1502 from the United States Department of Health and Human Services. Findings and conclusions presented in this report are those of the authors and do not necessarily represent the official position or policies of the United States Department of Health and Human Services.**

JAN 20 1999 NATIONAL CLEARINGHOUSE ON CHILD ABUSE AND NEGLECT INFORMATION

## TABLE OF CONTENTS

Page

SUMMARY ..... 1

INTRODUCTION ..... 5

EVALUATION DESIGN AND METHODS ..... 7
- Scales ..... 8

FINDINGS ..... 12
- Participant Characteristics ..... 12
- Pre- and Posttest Differences ..... 17
- Participant vs. Comparison Group ..... 18
- Within-Participant Group Differences ..... 21
  - Ethnicity ..... 21
  - Institution ..... 22
  - Father Status ..... 25

DISCUSSION AND CONCLUSIONS ..... 27

APPENDIX: DISCUSSION OF MISSING DATA ..... 28

## LIST OF TABLES


Table Page

1 YMAF Participants and Comparison Ward Characteristics .....13

2 Mean Pre- and Posttest Scores on Knowledge Scale.....17

3 Mean Pre- and Posttest Scores on Attitude Scale .....18

4 Mean Pre- and Posttest Scores on Esteem Scale.....18

5 Summary Statistics on Mean Differences Between Pre- and Posttest.....20

6 Summary Statistics on ANOVA models of Mean Scale Score (Ethnicity)...22

7 Summary Statistics on ANOVA models of Mean Scale Score (Institution) .23

8 Statistical Significance Levels on Mean Scale Score Differences .....25

9 Summary Statistics on Mean Scale Score Differences by Father Status.....26

## SUMMARY

Since 1993, the California Youth Authority (CYA) has been operating a parenting program called Young Men As Fathers (YMAF). This program, which is funded by a grant form the U.S. Department of Health and Human Services, was highlighted by Governor Wilson in his Focus on Fathers Summit in June 1995. It is designed for selected wards with the hope of making them good fathers by imparting them with the knowledge, attitudes, and self-esteem necessary for them to succeed.

YMAF is a 60-hour program offered currently to groups of 15 to 20 wards per class in four CYA institutions. The four institutions housing the YMAF wards are: DeWitt Nelson Training Center, El Paso De Robles School, Fred C. Nelles School, and Heman G. Stark Youth Training Center. The participation of wards in the YMAF program is continually expanding, and as of June 1996, two other CYA institutions were added: Karl Holton School and El Paso De Robles School. Governor Wilson has in fact directed CYA to expand its YMAF program to all Youth Authority institutions and camps and has allocated $2.7 million for its expansion.

Some of the program objectives include: strengthening the role of fathers in the families of participants, reducing child abuse and neglect, and improving the likelihood that the participants will provide emotional and financial support to their families. The objective of motivating these young fathers to provide financial support to their kids becomes critical because 75 percent of those YMAF participants who answered the question regarding whether they provided child support money to the mother of their child (children) indicated they did not send any support money.

This report is the first evaluation of the YMAF program. It is based on data collected between July 1993 and December 1995. During that period, 822 participants from four different CYA facilities were enrolled in the program; and they constituted the experimental group. The comparison group was formed by selecting wards form the Preston School of Industry and male wards from the Ventura School. Eighty-six percent of these participants graduated from the program. Additionally, data were also collected on 847 wards who served as the comparison group. These comparison wards were housed in two other CYA institutions.

The findings discussed in this report deal with the effectiveness of the YMAF program. Specific areas of interest include whether YMAF participants showed significant improvements in the areas of knowledge, attitude toward being a father, attitude toward being a father, and self-esteem that occurred between the beginning and the end of the classes, and between YMAF participants and the comparison wards. A summary of the findings shows:

- The YMAF program was effective in increasing the wards' knowledge of parenting skills and of the factual information (e.g., prenatal care, positive role modeling) taught in the class. This finding was reached by comparing the experimental group (comprised of the YMAF participants) and a comparison group of similar size on the basis of a Knowledge scale developed from pre- and posttest instruments. Possible scores for the Knowledge scale range from 0 to a maximum of 72. The experimental

group had a pretest mean score of 50.7 and a posttest mean score of 55.3. For the comparison group, the mean pretest score was 46.8 and the posttest score 46.5.

- YMAF participants did not show an improvement in their attitudes about being a parent between the beginning and the end of the YMAF classes, when compared to the matched comparison group on a similarly derived attitudinal scale. In fact both groups experienced significant negative changes in this scale. One possible explanation why YMAF participants displayed negative attitudes towards the classes (which presented information on the difficulties, challenges, and hard work associated with being a father) was that these classes made them realize that fatherhood was a difficult and challenging role to assume. Therefore, this interesting and unexpected finding can be interpreted in a positive manner since society does not want these kids to become fathers until they are ready to assume that role. Possible scores for the attitudinal scale range from 1 to a maximum of 9. The experimental group had a pretest mean score of 7.0 and a posttest mean score of 6.9. For the comparison group, the mean pretest score was 6.4 and the posttest score was 6.1.
- There was no evidence of significant improvement in the YMAF participants' self-esteem during the period before and after the classes, when compared to the comparison group on the basis of a similarly derived Esteem scale. Possible scores for the Esteem scale range from 10 to a maximum of 40. The experimental group had a pretest mean score of 32.5 and a posttest mean score of 33.3. For the comparison group, the mean pretest score was 31.9 and the posttest score 32.3.
- Within the experimental group, all ethnic groups performed well. In other words, the program appeared equally effective for the three collapsed ethnic groups. This

constitutes an important finding because the YMAF program was designed to be culturally sensitive.

Clearly, only the "real-life" application of this program--when the YMAF participants are released from institutions and start living their daily lives as fathers and interacting with the community--will ultimately indicate its success or failure. It is suggested here that a definite analysis of whether this program achieves its long-term goals of stemming delinquency and violence can be achieved only with a longitudinal research design that follows the YMAF participants over time.

Findings from this study are somewhat limited due to an unusually large amount of missing data. The problems of missing data and their impact on the evaluation are covered throughout this report, and more specifically in the Appendix.

# INTRODUCTION

As of June 26, 1996, the California Youth Authority (CYA) housed 9,546 male wards, 1,159 (or 12%) of whom have been identified as fathers.[1] All of these young men are faced with the prospect of being fathers if they are not already fathers; and most, if not all, are poorly prepared to assume the role.

It is well known that most CYA wards come from homes where parents and siblings have multiple social and personality problems. Many wards are from single-parent families where the absence of a father makes it even more unlikely that t hey will have an adequate parental role model; and, therefore, they will be ill prepared to assume a parental role themselves. Many Youth Authority wards simply have little concept of what it means and takes to be a parent.

Further, because the wards come from families where there is evidence of parental abuse and/or neglect, they will be very likely to treat their children in a manner similar to the way they were treated. If the wards of the Youth Authority were to continue to father children with the limited knowledge and skills of parenting they now possess, their children could become delinquents themselves and, thus, eventual wards of the Youth Authority.

As a result, the California Youth Authority developed an innovative program, the Young Men As Fathers (YMAF), to help male wards become fathers with the skills,

[1] This proportion of male fathers in CYA institutions is probably an underestimate because it is based on self-reports and incomplete court records. It does not include offenders who have lived or will live with mates who have children by other men and does not include those who are in a parenting role in their own families.

knowledge, attitude, and self-esteem necessary for them to assume their paternal role and to be a positive role model. The short-term goals of the program are to increase the parenting skills, knowledge, and self-esteem of the wards who participate in the YMAF program, while the long-term goals of the program are to reduce child maltreatment and delinquency in the future.

The program, funded by a grant from the U.S. Department of Health and Human Services, operates in the Institutions and Camps Branch. It was developed with input from CYA staff, outside parenting experts, and CYA wards who are fathers. This culturally sensitive program consists of 60-hour classes that are taught to groups of 15 to 20 male wards per class. The classes are taught from a standardized curriculum, but they are also designed to be highly interactive and to address the specific concerns and needs of the participants. The first classes of the YMAF program began in July 1993; and by December 1995, 822 male wards had enrolled in the program. Eighty-six percent of these wards graduated from the YMAF program.

## EVALUATION DESIGN AND METHODS

The purposes of this evaluation study are to: (1) investigate whether there were any significant changes in the YMAF program participants' knowledge of the information taught in class, attitudes toward being a father, and self-esteem that occurred between the beginning of class and the end of class; (2) show that any changes that occurred were due to the class and not some other factor; and (3) examine if the program was more effective for certain types of wards than others.

A test was administered to wards participating in YMAF both prior to the commencement of the class and at the end of the class. The test was designed to measure the effectiveness of the class with respect to knowledge of the information taught in class, the wards' attitude toward being a father, and the wards' self-esteem. The pre- and posttests were administered to both the experimental group, consisting of wards who participated in the program, and a comparison group, consisting of wards who did not participate in the program. The test was administered to the comparison group to show that any changes that occurred in the experimental group's knowledge, attitude, and self-esteem were due to participation in the YMAF program and not to some other factor.

The experimental group consisted of wards from the four institutions where the YMAF program was first implemented. The four institutions housing the YMAF wards are DeWitt Nelson Training Center in Stockton, El Paso De Robles School in Paso Robles, Fred C. Nelles School in Whittier, and Heman G. Stark Youth Training School (HGS/YTS) in Chino. All wards who participated in the YMAF program at these institutions were included in the experimental group.

The comparison group consisted of wards from the Preston School of Industry in Ione and the Ventura School in Ventura. The Preston School was specifically chosen to represent the northern California geographical region, and the Ventura School was chosen to represent the southern California geographical area. School officials at these two institutions chose one day, and one class during that day, during the week of April 11 through April 15, 1994, to administer the pretest to all wards who were in school that day. Then, ten weeks later, on the same weekday and during the same class period the posttest was administered. Staff at these institutions administered the posttest to all wards who were in school that day. This data collection strategy explains why there is a considerable amount of missing data on the comparison groups. In order for a comparison group ward to have valid data that could be analyzed, he had to be in school on both days and had to be in class during the hour the pre- and posttests were administered. If he was only present for one of the two days, then he would only be able to complete (theoretically) either the pretest or the posttest and, therefore, he would not be able to be included in the analyses conducted in this study.

**Scales**

Both the pre- and posttests consisted of two separate instruments. The first instrument was the curriculum test, and the second was the Esteem test. Three scales, the Knowledge, Attitudinal, and Esteem were developed from these two instruments.

The Knowledge and Attitudinal scales were constructed from the 70-item curriculum test. The Knowledge scale was constructed from 63 items that dealt with knowledge of the factual information presented in the YMAF program. This factual information included knowledge on prenatal care, infant and toddler stages, disciplinary

practices, medical concerns, parent-child bonding, and positive role modeling. The Knowledge scale ranged from a minimum score of 0 to a maximum score of 72. A score of zero meant that the ward did not answer any of the questions correctly, while a score of 72 meant that the ward obtained correct answers in all 63 questions. The maximum score is greater than the total number of questions because 5 questions had more than one correct answer.

The Attitudinal scale was constructed from the remaining 7 questions in the curriculum test that dealt specifically with the attitude of each ward toward being a father. The scale ranged from a minimum of 1 to a maximum of 9. Five of the 7 questions were worth one point each, while the remaining two were worth each up to four points. For those two four- point questions, a ward could obtain points for only one of them, not both of them; hence, the scale consisted of 6 items that were worth up to 9 points.

The Esteem scale was developed from the 10-item Esteem test. This test included several questions that were designed to measure the wards' self-esteem. The minimum possible score on this scale was 10, and the maximum possible score was 40. The items were all re-scaled prior to score calculations so that a high score represented high self-esteem and a low score represented low self-esteem.

A summary variable for each scale was constructed by subtracting the pretest score from the posttest scale.[2] For those wards who participated in the YMAF program, these variables represent the effect of participating in the program. More specifically, a positive value for these variables indicates that the program had a positive effect (i.e., the

[2] This variable is the "workhorse" of this paper and will be referred to throughout this paper as the scale score, even though it represents the difference between the pre- and posttests.

scale score increased), a score of zero indicates that the scores on the pretest and the posttest were the same, while a negative score indicates that the program had a negative effect (i.e., the scale score decreased). For those wards in the comparison group, the scale score differences are used to make sure that any changes that occur in the experimental group between the pretest and the posttest are due to participation in the class and not some other factor.

The methods of data analysis used in this study are bivariate statistical techniques.[3] More specifically, we use t-tests for dependent samples, t-tests for independent samples, and analysis of variance (ANOVA) models. We use three different methods since no one method by itself allows us to address all the objectives stated at the beginning of this evaluation. Each method allows us to address different substantive questions.

The t-tests for dependent samples allow us to investigate whether the program, for those wards who participated in it, increased or improved the wards' knowledge, attitudes, or self-esteem. For the comparison group, this allows us to investigate whether, for the Knowledge scale, there was a "learning effect" that occurred from taking the same tests twice. This type of t-test is used since we have a sample of cases with paired measurements on the same observation (i.e., pre- and posttest score). This type of t-test allows for controlling within-group variation, often known as error; and this increases the statistical power of the test.

[3] For a detailed explanation of missing data, and why simple bivariate statistical techniques were used rather than advanced multivariate statistical techniques, see the Appendix.

The t-test for independent samples is used to determine whether there is a difference in means (of the difference between the pre-and posttest) between the experimental and comparison group on the wards' change in their knowledge, attitudes, and self-esteem.

Analysis of variance (ANOVA) models and t-tests for independent samples is used to investigate if the program was more effective for certain types of wards than for others. The t-test for independent samples is used when there are only two groups of experimental wards to compare, and ANOVA models are be used when there are more than two groups of experimental wards to compare. ANOVA models were designed specifically to compare more than two means because of the inability of the t-test to compare more than two means.

# FINDINGS

In this evaluation, like in any other evaluation study, we are seeking an answer to the question, "is the program effective?" Clearly, the true effectiveness of this program can only be evaluated after the wards are released from CYA institutions and assume the role of a "father," but at this point in time we are simply asking whether the program was effective at changing the wards' knowledge, attitudes, and self-esteem. This chapter attempts to answer that question after it presents some descriptive findings.

## Participant Characteristics

Table 1 displays data for both the YMAF wards (the experimental group) and the comparison group wards. Of the 822 wards who enrolled in YMAF, 43% were Hispanic, 40% were African American, and 8% were white. Other minorities (Native Americans Asian, Filipino Pacific Islander, and other) accounted for the remaining 9%. The ethnicity of comparison group wards was 42% Hispanic, 33% African American, 13% white, and 12% "other."

The ethnic breakdown of the experimental and comparison groups in this study contrasts somewhat with the ethnic breakdown of the general CYA ward population. The general CYA population is composed of 46% Hispanic, 30% African American, 15% white, and 9% "other."

YMAF participants included in this analysis were housed in four CYA institutions: Heman G. Stark Youth Training School provided 47% of the experimental

Table 1

YMAF Participants and Comparison Group Ward Characteristics

| Category | Participants | | Comparison | |
|---|---|---|---|---|
| | Number | Percent | Number | Percent |
| Ethnicity | | | | |
| Hispanic | 352 | 43 | 353 | 42 |
| African American | 331 | 40 | 278 | 33 |
| White | 62 | 8 | 111 | 13 |
| Asian | 23 | 3 | 42 | 5 |
| Fil-Pacific Islander | 10 | 1 | 19 | 2 |
| Other | 44 | 5 | 35 | 5 |
| Total Cases | 822 | 100 | 847 | 100 |
| Mean Age | 19 | | 18 | |
| Marital Status | | | | |
| Not married | 378 | 66 | 521 | 88 |
| Would marry soon | 127 | 22 | 59 | 10 |
| Married | 42 | 7 | 7 | 1 |
| Divorced/separated | 24 | 5 | 7 | 1 |
| Total Valid Cases | 571 | 100 | 594 | 100 |
| Father Status | | | | |
| Have children | 371 | 65 | 133 | 22 |
| Wife/girlfriend pregnant | 34 | 6 | 22 | 4 |
| Plan fatherhood near future | 149 | 26 | 305 | 51 |
| No children now and in future | 19 | 3 | 138 | 23 |
| Total Valid Cases | 573 | 100 | 598 | 100 |
| Father's Marital Status | | | | |
| Married | 34 | 9 | 2 | 2 |
| Would marry soon | 94 | 26 | 28 | 21 |
| Not married | 220 | 60 | 97 | 74 |
| Divorced/separated | 19 | 5 | 4 | 3 |
| Total Valid Cases | 367 | 100 | 131 | 100 |
| YMAF Completion | | | | |
| Graduated | 707 | 86 | | |
| Did not graduate | 115 | 14 | | |
| Total Cases | 822 | 100 | | |

wards, DeWitt Nelson provided 23%, and Fred C. Nelles and El Paso de Robles each contributed 15%.

The comparison group consisted of 847 wards from two other institutions that were not yet participating in the YMAF program. The majority of them were incarcerated in the Ventura School (62%), while the rest was housed in the Preston School of Industry (38%).

YMAF participants' ages ranged from 14 to 25, with an average age of 19. This average age of YMAF participants matches that of the CYA population (19.1 years). About 69% of the YMAF participants were between the ages of 18 and 21.

Wards in the comparison group ranged from ages 14 to 19, and their average age was 18. Eighty-one percent of all comparison group wards were between 17 and 19 years old.

Data were also collected on the marital status of the experimental and comparison groups. Five hundred seventy-one of the 822 YMAF participants provided data on their marital status; 251 did not. Of the 571 valid cases, 66% were not married, 22% indicated they would be married soon, and 7% were already married. The remaining 5% of the participants were either divorced or separated, or in the process of being divorced or separated.

In the comparison group, 594 wards provided information on their marital status; 88% of them indicated they were not married, 10% said they had plans to get married soon, and only 1% was already married.

The wards were also asked to provide information regarding their father status. Of the 573 YMAF participants who provided valid data, 65% were currently fathers.

Further, nearly 60% of fathers were not married, over a fourth (26%) of the fathers indicated that they would be married in the near future, and only 9% of the fathers were married.

In the comparison group, 22% of the wards were fathers; of those wards who were fathers, 74% were not married, 21% indicated they would be married in a near future, and less than 2% of the fathers were married.

The vast majority of the YMAF participants (86%) graduated, and there were no significant differences in graduation rates by ethnicity.

Valid data on post-CYA living arrangements were available for 572 YMAF participants. Thirty-eight percent of them said they would live with their parents or stepparents, 35% would live with their wives or girlfriends, 12% with their relatives, and 8% would live by themselves. The remaining participants would either live with their friends or "other" people. A large majority of the 24% who said they would live with their wives or girlfriends will actually live with girlfriends since only 9% of the participants were married.

In the comparison group, 597 wards provided information on where they would live after they were paroled. The majority (57%) would live with parents or stepparents, 16% with relatives, and 14% said they would live with their wives or girlfriends. Almost all of these 14% would actually live with their girlfriends since only 1% of them were married.

Of the 371 wards enrolled in the YMAF program who indicated they were currently fathers, 302 provided information about whether they provided child support money to the mother of their child (children) while they were being incarcerated in CYA.

Of those who provided the information, 25% said they sent child support money, but 75% indicated they did not.

Three hundred twenty-nine fathers who participated in YMAF indicated their age, and the age of the mother of their child, when their first child was born. The mean father age was 16.8, and the mean mother age was 17.8. However, several mothers were over age 30 when the child was born, and these mothers are skewing right the distribution of the mothers' ages. Therefore, the median is a better measure of the mother's age at the child's birth. Using the median age, both the father and mother were 17 years old when their first child was born. Hence, the best conclusions that can be drawn are that: (1) the father ward and the mother of his first child usually had a similar age; and (2) if the father ward's age and mother's were different, then the father ward would be a little younger than the mother. This difference is surprising because for young people in general (i.e., not just CYA wards), the fathers are usually older than the mothers. However, for these CYA wards, it was not found that they were impregnating women that are younger than they are.

Three hundred fifty-eight wards from the experimental group provided information regarding their mothers' ages when they (i.e., the wards) were born; 462 did not. The mean age of the wards' mothers when they were born was nearly 23 years old. Over one-third of the wards' mothers were teenagers when they gave birth to the wards.

Finally, the age of the ward's mother when she gave birth to her first child was also provided. Only 348 YMAF participants gave this information. The mean age of the participants' mothers when they had their first child was 18.7 years, and nearly 68% of the wards' mothers were teenagers when they gave birth to their first child.

## Pre- and Posttest Differences

Table 2 presents the mean pre- and posttest scores on the Knowledge scale for both the experimental group (YMAF participants) and the comparison group. For the experimental group, there was a 4.61 point (or 9%) increase between the pre- and

Table 2

Mean Pre- and Posttest Scores on Knowledge Scale

| | Experimental | Comparison |
|---|---|---|
| Pretest mean | 50.71 | 46.85 |
| Posttest mean | 55.32 | 46.48 |
| Difference | 4.61 | -.37 |
| t-value | 14.95 | -.66 |
| Degrees of freedom | 523 | 245 |
| Significance | 0.000 | .508 |
| Cases included | 524 | 246 |
| Missing cases | 298 | 601 |

Posttest scores. This difference was highly significant with an associated significance level of .0000. More importantly, the comparison group's score on the posttest was lower than it was on the pretest by -.37 point; however, this difference was not significant.

Information on the mean pre- and posttest scores on the Attitude scale is presented in Table 3. Both the experimental and comparison groups had lower posttest scores than pretest scores, -.16 (or 2 %) and -.28 (or 4 %), respectively. Both these decreases were significant as well. In other words, both groups, at the time of the posttest, had more negative views about being a parent than they did at the time of the pretest. This is an interesting finding for which a possible hypothesis will be presented later, i.e.: why was there a negative change in the ward's attitude towards being a father after he attended classes designed to teach positive attitudes?

Table 3

Mean Pre- and Posttest Scores on Attitude Scale

| | Experimental | Comparison |
|---|---|---|
| Pretest mean | 7.04 | 6.38 |
| Posttest mean | 6.93 | 6.10 |
| Difference | -.16 | -.28 |
| t-value | -2.06 | -2.00 |
| Degrees of freedom | 521 | 242 |
| Significance | .040 | .047 |
| Cases included | 522 | 241 |
| Missing cases | 300 | 606 |

Summary information on the Esteem scale pre- and posttest scores is shown on Table 4. Both the experimental and comparison groups experienced increases in self-esteem, although only the experimental group's increase was statistically significant.

Table 4

Mean Pre- and Posttest Scores on Esteem Scale

| | Experimental | Comparison |
|---|---|---|
| Pretest mean | 32.46 | 31.95 |
| Posttest mean | 33.28 | 32.28 |
| Difference | .82 | .33 |
| t-value | 3.87 | .97 |
| Degrees of freedom | 365 | 230 |
| Significance | .000 | .330 |
| Cases included | 366 | 231 |
| Missing cases | 456 | 616 |

The experimental group's score increased from 32.46 to 33.28, which represents a 2.5% increase that is statistically significant at the .0000 level. The comparison group's score increased from 31.95 to 32.28 (+.33), but this difference wasn't significant at any level.

**Participant vs. Comparison Group**

With the exception of the comparison group on the Knowledge and Esteem scales, both the experimental and comparison groups experienced significant changes from the

pretest to the posttest on all the scales. While these findings are interesting by themselves, it is also useful to investigate whether there were any significant differences between the experimental and comparison groups. The prior findings only indicate whether the pre- and posttest scores in each group were significantly different and not whether the two groups were significantly different. It is possible that either: (1) there were significant changes in both the experimental and comparison groups, and either a significant or non-significant difference between the changes in the two groups; or (2) there was a significant difference in one of the group's pre- and posttest differences but not in the other group's difference, and either a significant or non-significant difference between the two groups changes. In other words, just because there is a significant difference within the groups does not mean that there is a significant difference between the two groups. The results presented here will indicate whether the mean differences in the pre- and post test scores of the experimental and comparison groups were significantly different. These findings will show whether the program was really effective at changing the participants' knowledge, attitudes, and self-esteem.

Table 5 presents the summary results on the tests of significance between the pre- and posttest mean differences between the experimental and comparison groups. There is a very significant difference in the mean change in the Knowledge scale score. The t-value and associated significance level for the difference between these two means are 8.37 and .0000, respectively. This information, coupled with the findings presented earlier, clearly indicate that the YMAF program is successful at imparting its participants with knowledge of positive parenting practices, skills, and knowledge, and that this change is arguably due to participation in the YMAF program

Table 5

Summary Statistics on Mean Differences Between Pre- and Posttest

| Scale | Experimental Mean | Comparison Group Mean | t-value | Degrees of Freedom | Significance |
|---|---|---|---|---|---|
| Knowledge | 4.61 | -.37 | 8.37 | 768 | .0000 |
| Attitude | -.16 | -.28 | .88 | 762 | .3800 |
| Esteem | .82 | .33 | 1.30 | 595 | .196 |

The results for the Attitudinal scale indicate that there wasn't a significant difference in the mean pre- and posttest difference. In other words, both groups experienced significant negative changes in the Attitudinal scale; but the significant change in the experimental group score was not significantly different from the comparison group's significant change. These results, given that both groups experienced significant negative decreases in their pre- and posttest scores, present a situation that is quite difficult to explain; but one possible hypothesis for these results is suggested.[4] One possible reason why the experimental group wards experienced a decrease in attitude was that the class, which presented information on the difficulties, challenges, and hard work associated with being a father, has made them see fatherhood as a difficult and challenging role. Similarly, for the comparison group wards, it is possible that the content of the pretest made them consider the hard work and difficulties associated with being a parent, and thus they had more negative views about fatherhood on the posttest. However, the fact that both groups experienced a significant decrease in their attitude towards being a father confounds any possible explanation about the program causing the change and highlights the fact that there could be some other

[4] The real cause of the decline can only be addressed in future research, possibly through interviews prior to and after participation in the YMAF program. One possible hypothesis for the decrease is presented here.

hypothesis, having nothing to do with the YMAF (e.g., an incarceration effect), that is causing the change in attitude in both groups.

The results in Table 5 also indicate that there wasn't a significant difference in the mean pre- and posttest change on the Esteem scale. The test of significance for these two means generates a t-value of 1.30, which has an associated significance level of .196. That is, even though the experimental group experienced a significant change in its pre- and posttest scores and the comparison group didn't, the mean pre- and posttest changes for the two groups were not significantly different from one another. This indicates that the mean difference between these two groups was not big enough to warrant a finding that indicates there was a significant difference between each group's mean Esteem scale score. This finding, similar to those in the Attitudinal scale, presents a situation that is not only difficult to explain, but one that is open to various interpretations as well. The results here only infer that the YMAF program does not appear to significantly increase the YMAF participants' self-esteem.

## Within-Participant Group Differences

This section examines whether the YMAF program was more effective for certain types of wards than for other types of wards. More specifically, it examines whether the program was more effective for wards who were: (1) of different ethnicity; (2) from different institutions; and (3) already or soon-to-be-fathers.

### Ethnicity

To conduct the analyses on ethnicity, 3 categories were collapsed into one: African American (n = 215), Hispanic (n = 215), and other (n = 94). The "other" category consists of whites, Native Americans, Asians, and Filipino and Pacific Islanders.

Table 6 displays summary statistics on the ANOVA models of the mean scale score pre-and posttest differences by these three ethnic categories. As seen in the table, there were no statistically significant differences between any of the mean scale score differences for any ethnic group.

Table 6

Summary Statistics on ANOVA Models of Mean Scale Score
Pre- and Posttest Differences by Ethnicity

| Scale | F-Ratio | Degrees of Freedom | Significance |
|---|---|---|---|
| Knowledge | .6770 | 2;521 | .51 |
| Attitude | .0586 | 2;519 | .94 |
| Esteem | .1954 | 2;363 | .82 |

The significance levels of all the F-ratios are far above the conventional levels of statistical significance (i.e., <.05). Substantively, this means that no ethnic group performed significantly better than any other on any of the pre- and posttest scale score differences. Thus, the program appears to be equally effective for all ethnicities. This is a very important finding since the curriculum was specifically developed to be culturally sensitive and, therefore, we should not have found any significant differences due to ethnicity.

**Institution**

As noted earlier, the wards who participated in the YMAF program and who are included in this analysis were housed in the Fred C. Nelles School, El Paso de Robles School, Heman G. Stark Youth Training School, and DeWitt Nelson Training Center. This section investigates whether the YMAF program was more effective at one (or more) institution(s) than any of the other institutions. Table 7 presents the summary statistics on the ANOVA models of mean scale score (of the difference between the pre-

and posttest) by institution. As indicated in the table, both the Knowledge and Attitude scales produce F-ratios that are statistically significant with p-values (i.e., significance levels) of .003 and .017, respectively; further, the Esteem scale nearly approaches significance with an significance level of .054.

Table 7

Summary Statistics on ANOVA Models of Mean Scale Score
Pre- and Posttest Differences by Institution

| Scale | F-Ratio | Degrees of Freedom | Significance |
|---|---|---|---|
| Knowledge | 4.5162 | 3;520 | .003 |
| Attitude | 3.3893 | 3;518 | .017 |
| Esteem | 2.5641 | 3;362 | .054 |

However, the fact that a significant F-ratio is found indicates only that it is unlikely that all population means are equal and, more importantly, does not indicate which means, if any, are significantly different from one another. The F-ratio is merely a preliminary test that indicates whether any further analysis is necessary. If the F-ratio is found to be non-significant, then the analysis is complete, and the means are satisfactorily concluded as being equal. This was the case in the prior section on ethnicity. When the F-ratio is significant, then further analysis is necessary, and this analysis involves the use of multiple control procedures. Multiple comparison procedures are used to find means with significant differences[5]. Further, multiple comparison procedures prevent one from calling too many means significantly different because they adjust the significance levels according to how many comparisons are being made (i.e., the more comparisons, the larger the difference that is required to call two means significantly different).

Table 8 presents the results of the Tukey's Honest Significant Difference for Unequal Size Groups multiple comparison procedure for each of the three scales. This procedure, developed by Spjotvoll and Stoline (1973), is a modified version of Tukey's Honest Significant Difference procedure that makes adjustments when the groups are of unequal sizes. This procedure, as opposed to other available procedures such as Tukey's Honest Significant Difference Test, Newman-Keuls Test, and Scheffe Test, was specifically chosen since it takes unequal group sizes into account when adjusting the significance levels[6]. The numbers in the tables are the levels of significance between the two respective institution means. Any probability less than .05 indicates that there is a statistically significant difference between the means of the two institutions. In this case, however, no means on any of the scales were found to be significantly different after adjusting for the number of comparisons being made and the unequal group sizes. In fact, the institution means with the smallest significance level (and hence the highest probability of a significant difference) are the Heman G. Stark Youth Training School and El Paso de Robles School means on the Knowledge scale, and the Heman G. Stark Youth Training School and DeWitt Nelson means on the Attitude scale, each with an observed significance level of .10. This level of significance is not

[5] The mathematical foundations and statistical properties of multiple comparison procedures are complex and far beyond the scope of this evaluation, and the reader is referred to Neter et al. (1996) for a thorough introduction to them.

[6] The Scheffe Test and the Newman-Keuls Test were also applied to the data, and each of these methods generated the same substantive conclusions as the Tukey's Honest Significant Difference Test we used.

Table 8

Statistical Significance Levels on Mean Scale Score Differences
by Institution

| Scale | Institutions | | | |
|---|---|---|---|---|
| | Nelles | El Paso de Robles | DeWitt Nelson | HGS/YTS |
| Knowledge | | | | |
| Nelles | - | - | - | - |
| El Paso de Robles | .96 | - | - | - |
| DeWitt Nelson | .45 | .35 | - | - |
| HGS/YTS | .25 | .10 | .76 | - |
| Attitude | | | | |
| Nelles | - | - | - | - |
| El Paso de Robles | .95 | - | - | - |
| DeWitt Nelson | .97 | .98 | - | - |
| HGS/YTS | .54 | .52 | .10 | - |
| Esteem | | | | |
| Nelles | - | - | - | - |
| El Paso de Robles | .82 | - | - | - |
| DeWitt Nelson | .99 | .69 | - | - |
| HGS/YTS | .77 | .99 | .13 | - |

statistically significant at conventional levels of significance. As a result, it can be inferred that there is no significant difference between the mean pre- and posttest differences on any of the scale scores; and, therefore, the program appears to be equally successful in all four participating institutions. This is a reflection of good consistent training of the instructors on the curriculum.

**Father Status**

As indicated earlier, 65% of the YMAF participants were already fathers at the time the program began, 6% had girlfriends or wives who were pregnant, and the other 29% currently were neither fathers nor had girlfriends or wives who were pregnant. The question of whether the YMAF program was more effective for those wards who were either already fathers or whose girlfriends or wives were pregnant than for those who were neither of those will now be examined. It is possible that those wards who were

either fathers or soon-to-be fathers had more incentive or motivation to learn and, therefore, may have had significantly different mean scale score differences in comparison with the other "non-father" wards. Since there are only two groups in this analysis, the t-test for independent samples is used to examine whether there are any significant differences between these two groups.

Table 9

Summary Statistics on Mean Scale Score Differences
by Father Status

| Scale | t-value | Degrees of Freedom | Significance |
|---|---|---|---|
| Knowledge | 1.29 | 486 | .199 |
| Attitude | .96 | 464 | .338 |
| Esteem | .82 | 358 | .412 |

Table 9 presents the summary statistics on the mean scale score differences. None of the t-values for any of the scales is statistically significant. Therefore, one can conclude that the program was no more effective for fathers/soon-to-be fathers than for those wards who were neither fathers nor soon-to-be fathers. The mean scale differences (between the pre- and posttest scores) appear to be equal for both the wards who were "fathers" and the wards who were "non-fathers."

## DISCUSSION AND CONCLUSIONS

The analyses carried out in this study indicate that (1) the YMAF program is effective at increasing the wards' knowledge of parenting skills and of the factual information (e.g., prenatal care, positive role modeling) that is taught in the class; (2) no significant differences were found between the experimental and comparison groups on the Esteem and Attitude scales; (3) ethnicity, institutional location, and paternity were not significantly related to the measurable outcomes.

These findings should be considered within the context of the limitations of this evaluation: missing data, non-random samples, and the inability to statistically control for any differences between the experimental and comparison groups. All of these limitations could be distorting the findings made in this report. These limitations also suggest that additional research be conducted to determine effectiveness of the YMAF program. Governor Wilson has ordered, via Item 7 of Executive Order W-125-95, that the YMAF program be expanded to all of the CYA institutions and camps; and, therefore, more wards will be going through the YMAF program than ever. Future research is necessary so that program areas can be fine-tuned, expanded, or eliminated to ensure that the program is as cost-effective as possible. Finally, it should be emphasized that the real effectiveness of the program can only be evaluated when the wards assume their roles of fathers, once they are released from CYA institutions or discharged from parole. A definitive analysis of whether this program achieves its long-term goals of stemming delinquency and the inter-generational cycle of violence can be achieved only with a longitudinal research design that follows the YMAF participants over time.

## APPENDIX: DISCUSSION OF MISSING DATA

This appendix includes a discussion about missing data, the effects of missing data on our choice of statistical techniques, and the effects of missing data on the generalizability of the conclusions of this evaluation.

The large amount of missing data encountered during the course of this evaluation hindered the analysis to a large extent. There are several reasons why so much data are missing: (1) wards dropped out of the program; (2) the institutions failed to either administer or send the pretest and/or posttest to the Research Division; (3) the curriculum test was not finalized prior to the commencement of the first class (starting in July 1993; n = 134) and, therefore, was not administered to the wards in that class; and (4) wards failed to comprehensively complete the pretest and/or posttest. Of those reasons, the primary reason for missing data was the failure of the wards to complete the tests entirely. However, it should be noted that as time went on, the amount of missing data decreased. That is, proportionately speaking, there were less missing data found in the later classes than in the earlier classes; but the later classes were still, on average, missing about 15% of the data needed to perform the evaluation analyses. The Research Division made all possible attempts to collect all missing data (e.g., accessing OBITS, calls to institutions), but for many cases and for certain variables, this information could not be found or retrieved.

Modern statistical software programs, such as SPSS, have several ways of handling missing data when calculating estimates. One way is pair-wise deletion, which only deletes cases from the estimation of the effects of variables for which there are

missing data, but each case is included in the estimation of the effects of all variables for which they have valid data. Another way is case-wise deletion, which deletes any case that has a missing data for *any* variable in the analysis. In other words, analyses using the case-wise deletion option only include cases in the analysis if they are missing no data on any of the variables in the analysis. A third option is to substitute the mean value of the variable for all cases that are missing data on that variable.

The choice between these three methods of dealing with missing data makes little difference when the amount of missing data is small. However, there are large discrepancies between the methods when the amount of missing data is large. The main problem with the case-wise deletion is that when the missing data are randomly distributed across cases, one can end up with very few or no valid cases, since each case would hypothetically be missing data on at least one variable. That is, the sample size can essentially decrease to zero.

The problem with the pair-wise deletion method is that it fails to produce a "true correlation matrix," which is a matrix where all correlations between each of the variables are all calculated using the same cases. This becomes a major problem because the correlation matrix is used in nearly all the estimation procedures of advanced statistical techniques. In a multivariate analysis, the pair-wise deletion method allows the sample size to vary for each of the variables in the analysis, which can be very questionable and risky. One would not want the effect of one variable to be calculated on 2,000 of cases and another to be calculated on only 45 of them, and then compare which of these variables has the greatest effect on the dependent variable. However, the pair-wise deletion method presents no real biases as long as the amount of missing data is less than

10% and as long as the missing data are randomly distributed across cases, not systematically distributed among them.

The problems with the mean substitution method are that: (1) it artificially decreases the variance of the data for that variable, and this decrease is directly proportional to the amount of missing data, that is, the more data that are missing, the more data will consist of artificially created average scores; (2) if the mean is not a good representation of that variable (i.e., if the variable is not unimodally distributed), then substituting the mean can create severe bias; and (3) because this method substitutes artificially created average scores, it may significantly decrease the correlation between any variables that have a strong correlation.

The prior discussion on missing data and how statistical software programs deal with them was an extremely important concept for choosing an appropriate statistical technique for the analyses in this evaluation. As stated earlier, the purpose of using a comparison group is to ensure that any differences that occur between the pre- and posttest are due to participating in the program and not to some other factor. There are two primary ways an analysis using an experimental and comparison group can be conducted. The first one is to construct a comparison group by matching each experimental group case with an equivalent comparison group case. The matched comparison group should resemble the experimental group on all relevant variables that could exert influence on the results. Therefore, any difference between the experimental and comparison groups can be argued to be due to the program and not some other factor. In order to ensure that proper matching occurs, the sample size of possible comparison group cases usually needs to be more than five times the size of the experimental group.

Further, it becomes extremely difficult to construct a matched comparison group when cases being matched contain variables with large amounts of missing data.

The second way to conduct an experimental and comparison group evaluation study is to statistically control for any differences between the experimental and comparison groups. Advanced multivariate statistical techniques, such as multiple regression, loglinear modeling, and logistic regression, allow researchers to compare the outcomes of the participants in the program with the outcomes of the non-participants, while statistically controlling for any (measured) differences between the experimental and comparison groups. This allows researchers to remove possible selection bias (e.g., being a father in the present study) that occurs when individuals are not randomly assigned to either the comparison or experimental group.

For the current analysis, it was possible not to use either of these techniques to control for the differences between the two groups. This fact needs to be remembered when considering the generalizability of the results presented in this study. Matched comparison groups could not be constructed for several reasons.

First, the comparison group consisted of 847 subjects, while the experimental group consisted of 822 subjects. In other words, the two groups would already have to have been equivalent on all the relevant control variables for them to be matched effectively. Secondly, there were substantial amounts of missing data on many of the relevant variables used to match cases.

Similarly, it was impossible to statistically control (in the multivariate sense of statistical control) for the differences between the two groups because of the large amount of missing data. There was no way to adequately deal with the amount of missing data.

If case-wise deletion were used in a multiple regression analysis that controlled for all the relevant variables, the sample size would shrink from 1,669 to 157. This occurs because most cases were missing data on at least one crucial variable; and, therefore, they would be eliminated from the analysis entirely. Any conclusions that could be made from such an analysis would be neither reliable nor valid. Hence, the case-wise deletion method could not be used.

If pair-wise deletion was used in a multiple regression that controlled for all the relevant variables, the multiple regression model would have been estimated on a sample of 770. Remembering that the original sample size was 1,669 and that most cases were missing data on only one of the control variables, the reason the sample size decreases to 770 in pair-wise deletion is because many cases were missing values on the dependent variables as well. If a case is missing datum on the dependent variable, it is excluded from the analysis altogether in the pair-wise deletion method, as well as in the case-wise deletion method and the mean substitution method. Necessarily, this means the largest sample size, which could be used in any of the analyses, would be 770, not 1,669. However, the main problem with the pair-wise deletion method was that it produced a correlation matrix which was far different from the correlation matrix produced with the case-wise deletion method. If the correlation matrix produced by the pair-wise deletion method does not contain any systematic bias, than the case-wise deletion method will produce a correlation matrix that is similar to the pair-wise matrix. If they are different, then there is good reason to believe there is systematic bias present in the correlation matrix of the pair-wise deletion method. In this case, this systematic bias is due to the fact that the missing data are not only randomly distributed but are also systematically

distributed as well (i.e., certain cases or groups of cases are more likely to be missing data on a particular variable). Therefore, the correlations that appear in the correlation matrix of the pair-wise deletion method are all calculated on different subsets of cases. This method of dealing with missing data produced a correlation matrix which was not a "true correlation matrix," which advanced statistical techniques require to produce valid and reliable results. Therefore, the pair-wise deletion method could not be used.

Thirdly, there was simply too much missing data to justify substituting the mean of a variable in its place. Some variables were missing data for up to 70% of the cases, and to substitute the mean for such variables would not be a very acceptable practice and would produce biased, invalid, and unreliable results. Hence, the mean substitution method could not be used.

As a result of all of the problems associated with missing data and the advanced statistical techniques that would have allowed for controlling for any differences between the experimental and comparison groups, the decision was to use of several bivariate statistical techniques that would be more practical and would generate more valid and reliable results. Conducting the analyses with bivariate statistical techniques allowed for avoiding many of the problems that occur with missing data and advanced multivariate statistical techniques, although they limit the generalizability of the results. Because we could not control for the differences between the experimental and comparison groups, the analyses that were conducted here made the implicit assumption that the cause for the differences between the two groups were due to the YMAF program and not some other factor(s). This left the authors with no other option but to proceed with the analyses based on that assumption. Obviously, this limits the generalizability, reliability, and

validity of the results, but the authors believed this method of analysis produced more generalizable, valid, and reliable results than would have been produced by the more advanced statistical techniques.

STATE OF CALIFORNIA--YOUTH AND ADULT CORRECTIONAL AGENCY

GRAY DAVIS, GOVERNOR

**DEPARTMENT OF THE YOUTH AUTHORITY**
4241 Williamsbourgh Drive, Sacramento, California 95823
Telephone (916) 262-1392

TTY (800) 555-6469 - California Relay Service (800) 735-2922
Internet Address - www.cya.ca.gov

January 6, 1999

Joyce Zarrinnahad
National Clearinghouse on
Child Abuse and Neglect
330 C St., SW
Washington, DC 20447

Attention: Mr. Christopher Ellis

Dear Mr. Ellis,

Enclosed is the Final Report for the "Young Men as Fathers Program" Grant Number 90CA1502. Although we did submit our progress reports, fiscal reports, and other documents on time, this Final Report is actually a research document, which was completed long after the program ended.

We apologize for the delay in getting this report to you. It was reviewed by a number of people and took longer than we expected to publish. It was an unintended oversight that we did not forward a copy to you.

Thank you for reminding us of this obligation. You might also want to note that as a result of this "seed money," the State of California provided $3M. to the counties to replicate the program in 1996-97 and for 1998, we are allocated another $1M. All of the Youth Authority sites, which had the program, are ongoing and in fact, all of our 16 facilities and camps now have the Young Men as Fathers program.

This program was highly successful and we are hopeful that in the future generations our efforts will be even more obvious—in that the children of young offenders do not end up hurting other people and being incarcerated.

Please let me know if you need additional information.

Sincerely,

Sharon J. English
Deputy Director
Office of Prevention and Victim Services

Enclosure

cc: Frank Alarcon
Elaine Duxbury
Jeff Lara
Kip Lowe