

I. INTRODUCTION

The third year of program evaluation of the Mentors in Violence Prevention (MVP) Program took place during the 2001-2002 academic year. In this report, the first and second years of evaluation activity are summarized briefly, and the goals of the third and final year of evaluation are outlined. Year 3 evaluation results are presented, and findings concerning the impact and effectiveness of the MVP program are discussed in the context of the multi-year evaluation strategy that was employed.

1.1 The First Two Evaluation Studies of MVP

The multi-year, multi-method approach to evaluating the MVP Program is a strong one. Unlike many evaluations of adolescent gender violence prevention programs reported in the literature, the evaluation of the MVP Program spans three successive academic years and includes a significant qualitative element that helps both to explain the need for and the meaning of such programming in students' lives as well as to ensure that student voices are well represented in the study. Moreover, the quantitative component of the evaluation included a survey designed specifically to assess the impact of participation in the MVP Program. The survey instrument was administered at both pre- and post-test during all three years of evaluation, affording greater opportunity to make comparisons and greater confidence in the findings that were replicated across multiple years of evaluation activity.

1.2 Goals of the Year 3 MVP Evaluation Study

The third year of evaluation of the MVP Program had two main purposes:

1. To replicate the quasi-experimental, pre-test/post-test design with comparison group;
2. To evaluate MVP Program outcomes, paying attention to gender differences, including:
 - a. levels of student knowledge and awareness;
 - b. student attitudes concerning gender violence and one's own ability to be an active bystander;
 - c. student satisfaction with the MVP Program.

Year 3 survey results are presented in the pages that follow.

II. SURVEY EVALUATION METHODS

A quasi-experimental, pre-test/post-test survey design with comparison group was used to evaluate the impact of the MVP Program on student knowledge and attitudes. The *MVP Survey* (see Appendix C) was administered for these purposes. The survey instrument was designed specifically to evaluate the MVP training curriculum. Previous validity and reliability testing have demonstrated that the *MVP Survey* is a sound measurement tool. Qualitative data were elicited from an open-ended questionnaire gauging student satisfaction with the MVP Program.

2.1 Sampling

Throughout the 2001-2002 academic year, MVP trained students at eleven different high schools. All students from the 11 schools who participated in the initial awareness-raising training were asked to participate in the evaluation by filling out the survey instrument. Thus, no sampling criteria were employed for the population of students who were in the MVP training.

Students from three schools were selected to constitute a comparison group. Random sampling was not employed. Instead, these schools were chosen as a convenience sample based on their willingness to serve as comparison sites for a year in exchange for receiving MVP's services during the 2002-2003 academic year. Despite this constraint, efforts were made to select comparison sites that were comparable to treatment sites in terms of student demographics. School personnel at the comparison sites also were asked to select student participants in the same manner (i.e., using the same leadership criteria) they would employ as if the selected students were receiving training. Students at these sites did not receive any information about or training from MVP during the 2001-2002 academic year.

2.2 Survey Administration Procedures

Pre-test surveys were administered to students on-site in a group administration setting. Student pre-testing took place prior to any exposure to the MVP Program. At treatment sites, pre-test survey administration took place on a rolling basis, as this is the manner in which MVP provides services to schools. For example, some schools began MVP training in the Fall of 2001 while others did not begin until the Spring of 2002. Typically, students in the treatment group filled out the pre-test survey immediately before their first training session. Post-test surveys were administered to students in the treatment group approximately two weeks after they completed the MVP awareness-raising training. Again, testing was on-site, in a group setting. Because of differences pertaining to training schedules at individual sites, the time in between pre- and post-testing at treatment sites varied from school to school and ranged from less than two months to almost five months.

Pre-test surveys also were administered to students at the comparison sites on a rolling basis, and the amount of time between pre- and post-test administration varied at these sites as well. Post-test surveys were administered at the comparison sites between one and four months after pre-testing. Pre- and post-test surveys were administered on-site in a group setting. This testing condition was similar to the treatment condition and should pose little, if any, threat to validity.

Prior to any testing, all students were verbally informed of the purpose of the MVP evaluation and were ensured that their responses to the *MVP Survey* would remain confidential. Students were told that their participation was voluntary, that they could withdraw from the study at any time without penalty, and that they did not have to answer any questions that made them feel uncomfortable.

2.3 Response

Two schools in the treatment group dropped out of the MVP Program mid-way through training due an inability to organize effectively in order to undergo MVP training. These schools were not included in the analysis; thus, nine schools participated in the Year 2 evaluation study of the MVP Program. The treatment group was comprised of 244 students from these nine schools. Fifty-seven percent of

respondents from the treatment group were girls, while 43 percent were boys. One of the comparison group sites dropped out of the study, so students from only two schools formed the comparison group for this study. From the comparison schools, 65 students participated in the study. Forty-six percent of comparison group students were female, and 54 percent were male. Chi square tests showed that there were no significant gender differences between treatment and control groups. The mean age of students in both the treatment and comparison groups was 16 years.

Thirty-seven percent of student respondents from the treatment group were Caucasian. Twenty-seven percent were African American. Fourteen percent were Latino, 7 percent were Asian, and nearly 15 percent identified as coming from a mixed ethnic heritage. In the comparison group, a majority of students (53 percent) were Caucasian. Twenty-one percent of comparison students said they came from a mixed ethnic background. Fourteen percent of students in the comparison group identified as African American, 11 percent as Asian American, and less than 2 percent of students in the comparison group were Latino. Chi square tests showed that there were significant differences between the treatment and comparison groups in terms of race and ethnicity.

III. YEAR 3 SURVEY RESULTS

This section details the analysis and results of the third year survey data. First, changes in student knowledge and awareness levels are described. Next, changes in student attitudes regarding gender violence are presented, followed by a discussion of changes in student attitudes regarding their ability to prevent and/or intervene in situations involving gender violence. Results from the fourth section of the *MVP Survey* are then described.

3.1 Knowledge and Awareness

Five survey questions were designed to measure the level of student knowledge and awareness regarding gender violence. The five questions are factual items based on information that is underscored during MVP training and are found in Part I of the *MVP Survey*. For these questions, students are asked to respond by indicating whether they believe the statement is true or false. A third "I don't know" option has been included to minimize guessing and non-response.

In order to analyze the nominal data in the knowledge section of the survey, cross-tabulations and Pearson's Chi Square tests were conducted for each of the five variables (KV1-KV5) using the statistical software package, SPSS (Statistical Package for the Social Sciences). Each of the five variables was analyzed separately, as the items do not form a scale. Pearson's Chi Square tests were performed at the $\alpha = .05$ and $\alpha = .10$ levels. The purpose of these tests is to demonstrate whether the change in knowledge from the pre-test and the post-test differs between the treatment and comparison groups. **Statistically significant treatment effects were noted for several of the knowledge variables.** Group differences are presented descriptively below.

*KV1: “According to MA law, it may be considered rape if a man has sex with a woman who is under the influence of alcohol or other drugs.” (The correct answer to this question is **true**.)*

For this first knowledge variable (KV1), 73 (142/194) percent of students in the treatment group (i.e., those who received the MVP Program) answered correctly at pre-test. The percentage of treatment students responding correctly to this item at post-test rose to 97 (188/194) percent, compared to 84 (52/62) percent of students in the comparison group. Strong treatment effects were found for this variable. For example, while 100 (18/18) percent of MVP students who answered false on the pre-test changed their answer to the correct “true” response at post-test, only 33 (1/3) percent of comparison students did so. This difference between treatment and comparison groups was statistically significant ($p = .000$). Additionally, 85 percent of MVP students changed their pre-test “I don’t know” response to the correct “true” response at post-test. This was significantly different ($p = .030$) from the 50 percent of students in the comparison group who did so. As in the Year 2 evaluation, findings from the Year 3 survey suggest a rather powerful treatment effect for this variable – students exposed to MVP training have greater knowledge regarding the state rape laws in Massachusetts.

*KV2: “As long as you are just joking around, what you say or do to someone cannot be considered sexual harassment.” (The correct answer to this question is **false**.)*

As was the case in MVP’s Year 2 evaluation, most students knew the correct answer to this question at pre-test (92 percent of both treatment (178/194) and comparison (55/60) group participants answered correctly at pre-test). A majority of students in the treatment and comparison groups (96 and 90 percent, respectively) knew the correct answer at post-test. Though a greater percentage of MVP students knew the correct answer to this question at post-test, statistically significant differences were not realized for this variable during Year 3. This stands out as the single variable for which there was little difference between treatment and comparison groups and little improvement from pre- to post-test among the MVP students, likely because such a high number of students knew the correct answer at pre-test.

*KV3: “People on TV and in the movies influence the ways that we act as men and women.” (The correct answer to this question is **true**.)*

Roughly the same number of treatment and comparison group students knew the correct answer to this question at pre-test (73 percent of treatment group students and 74 percent of comparison group students). At post-test, 81 percent of student who received the MPV Program answered correctly versus just 70 percent of comparison group students. As a whole, while treatment group students improved on this variable over time, comparison group students exhibited a lower level of knowledge at post-test than at pre-test. Ninety-one percent of MVP students who answered correctly at pre-test maintained their correct answer at post-test as compared with 76 percent of students in the treatment group. This difference was statistically significant ($p = .01$).

KV4: “Men perpetrate (commit) over 90% of violent crime in the U.S.” (The correct answer to this question is **true**.)

Treatment effects on this remain strong in Year 3, replicating the results of Year 2 testing. Only 28 percent of students in the MVP Program answered correctly at pre-test. This percentage rose to 85 percent at post-test. Conversely, while 39 percent of comparison group students answered correctly at pre-test, this percentage dropped to 31 percent at post-test. In fact, nearly 40 percent of those in the comparison group who answered correctly at pre-test changed to an incorrect answer at post-test. This was a significant difference ($p = .025$) from the 85 percent of MVP students who maintained their correct answer upon Program completion. Similarly, the 67 percent of MVP participants who answered “false” on the pre-test but chose the correct “true” response at post-test was significantly different ($p = .000$) from the 11 percent of comparison group respondents who did so. Lastly, 73 percent of students in the treatment group correctly changed their pre-test “I don’t know” response to “true” upon Program completion. Only 11 percent of comparison students did so and, again, this difference between the two groups was strongly significant ($p = .000$).

KV5: “In the U.S., a man physically abuses a woman every 9 to 18 seconds.” (The correct answer to this question is **true**.)

In the Year 3 study, important treatment effects were again found for this fifth knowledge variable. The percentage of MVP students who knew the correct answer at post-test jumped to 78 percent from just 35 percent at pre-test. In contrast, while 40 percent of comparison group students answered correctly on the pre-test, only 32 percent chose the correct answer at post-test. Moreover, while 84 percent of MVP students maintained the correct answer from pre- to post-test, only 56 percent of students in the comparison group did so. This difference between groups was statistically significant ($p = .010$). Additional group differences were seen among students who incorrectly answered “false” during pre-testing. While 88 percent of students who received MVP training correctly changed their answer to “true” at post-test, just one student (14 percent) among the comparison group did so. This difference was strongly significant ($p = .001$). Similarly strong effects were seen between students who answered “I don’t know” at pre-test but correctly answered “true” at post-test. The 72 percent of treatment group students who correctly changed their answer was statistically different ($p = .000$) from the 17 percent of comparison group students who did so.

In sum, participation in the MVP Program seems to heighten the level of knowledge and understanding that students who are exposed to training have regarding gender violence. Important differences between MVP and comparison group students were found for the five knowledge variables on the Year 3 survey. In every instance, a greater percentage of MVP participants answered correctly at post-test. Further, statistically significant differences emerged between the treatment and comparison groups in four out of the five areas that were tested. These findings replicate Year 1 and Year 2 MVP evaluation results, providing ever stronger evidence of MVP’s effectiveness at increasing student knowledge and awareness. Results from knowledge tests across the multi-year evaluation of the MVP Program are found in Table 3.1 below.

TABLE 3.1

Significant Results on Knowledge Variables by Study Year¹

	Year 2	Year 3
KV1	Strong treatment effects	Strong treatment effects
KV2	Weak Treatment effects	No treatment effects
KV3	Significant Treatment effects	Significant Treatment effects
KV4	Strong treatment effects	Strong treatment effects
KV5	Strong treatment effects	Strong treatment effects

3.2 Attitudes Regarding Gender Violence

Part II of the *MVP Survey* was designed to measure student attitudes toward male violence against women as conceptualized by the MVP Program and consistent with the literature on adolescent gender violence. In the MVP context specifically, indicators of gender violence can range from telling jokes that objectify women and possessing stereotypical ideas of gender roles to rape and battering.

The sixteen items found on Part II of the survey form a unidimensional scale (the “AV Scale”). Examples of items found on this scale include, “It is harmless to tell dirty jokes about women” and “If a guy forces his girlfriend to have sex with him when she doesn’t want to, it is rape.” Several items are reverse-worded to reduce systematic error. Statements are followed by a five-point, Likert-type scale response set ranging from “Strongly Disagree” to “Strongly Agree.” **On this scale, higher scores indicate ignorant or sexist attitudes about gender violence.**

Survey responses were coded, entered into a database, and analyzed using SPSS. The reliability of this scale was analyzed using Cronbach’s alpha². For the pre measure, $\alpha = .76$, and for the post measure $\alpha = .85$, indicating that the scale is highly reliable and will produce consistent results over time. To help assess the construct validity of the scale, Principal Component Analysis was implemented. Construct validity helps to ensure that one is measuring what one intends to measure. Factor loadings for the AV Scale demonstrate that it is reasonable to retain all 16 items and to maintain the scale without dividing it into subscales; however, arguments could be offered to make some changes to the AV Scale.

The principal component scale explains 32 percent of the variation among the 16 original AV Scale items in the post-test and roughly 25 percent of the variation among those items in the pre-test. While

¹ Year 1 results are not included in Table 3.1, because 3 of the 5 knowledge questions were changed beginning with implementation of the Year 2 survey.

² Past efforts at assessing the validity and reliability of the *MVP Survey* are described in more detail elsewhere. Findings were similar across MVP Evaluation studies. For more information, contact the researcher through the MVP Program.

the “explained” variations are somewhat low, an averaged scale score that used the mean values of the 16 scale items (computed separately for the pre- and post-test) was created to ensure that the results were robust to changes in scaling procedures.

Given the interval nature of the dependent measure, multivariate regression analysis techniques were employed to study group effects. Two regression models were run for the AV Scale using the newly created variable that is the average across the 16 measures (described above). One regression was run without including any independent variables. Another was run controlling for several independent variables, including the Pre AV mean score, group assignment (treatment vs. comparison), gender, race, age, and survivor status.

Strong treatment effects were found using both regression models. Being in the MVP Program is associated with a lower post-test score (recall that higher scores indicate sexist attitudes) than being in the comparison group, and the difference is significant ($p = .000$). This result mirrors the findings from the first and second years of evaluation of the MVP Program.

Table 3.2.1		
AV SCALE		
Mean Score by Group		
<i>(Higher mean scores indicate ignorant or sexist attitudes about gender violence)</i>		
	Treatment Group Mean	Comparison Group Mean
Pre-Test	2.02	2.12
Post-Test	1.78	2.23
Change from Pre- to Post-Test	-0.24	0.11

It is also helpful to discuss the impact that the MVP Program had on participants’ attitudes toward gender violence more descriptively by comparing the mean scores of the two groups (see Table 3.2.1). Notice that at pre-test, treatment and comparison group mean scores were similar. The treatment group mean was 2.02, the comparison group mean was 2.12, and these were not significantly different ($p = .102$). At post-test, the mean score of MVP students had dropped from 2.02 to 1.78, while the mean score for the comparison group had actually increased by .11 points to 2.23. The difference between the two groups at post-test was, again, statistically significant ($p = .000$). Upon Program completion, attitudes among MVP participants are shown to have improved, while the attitudes among comparison group students actually got worse.

While relatively small sample sizes limit the ability to statistically gauge differences among subgroups, it is important to try to understand such effects. Findings from the Year 3 study point once again to differences between the young women and men who receive the MVP Program. Girls begin and end the MVP Program with more desirable attitudes than do boys, and the difference is significant ($p = .024$). Limited power makes it difficult to find both direct and indirect effects using the regression

models above; consequently, gender differences are presented here descriptively in Table 3.2.2. Among students who received MVP training, girls had lower mean scores than boys at both pre- and post-test. The difference is somewhat pronounced at both testing instances, indicating that girls begin and end MVP Program training with more desirable attitudes than their male counterparts. This result is consistent with the findings from both of the previous years of survey implementation. In addition, while mean scores on the AV Scale changed significantly for both girls and boys from pre- to post-test, the change is slightly more pronounced for boys. This, too, is consistent with findings from the two previous evaluation studies, indicating that the MVP Program has a slightly greater impact on boys than girls in terms of this construct. It is meaningful that this result is replicated across studies. Without the benefit of larger sample sizes and specific statistical testing, however, definitive conclusions cannot be drawn.

In the Year 3 study, for the first time, program effects were also found in terms of differences between African American students and Caucasian students. Specifically in Year 3, being African American was associated with a higher post-test score on the AV Scale, and the difference is significant ($p = .020$). This indicates that MVP may have been more effective for Caucasian students than for African American students; however, this finding is difficult to interpret given that there were significant differences in the racial/ethnic composition of the treatment and comparison groups.

Table 3.2.2				
AV				
Mean Scores by Gender and Group				
<i>(Higher mean scores indicate ignorant or sexist attitudes about gender)</i>				
	Treatment Group		Comparison Group	
	Girls	Boys	Girls	Boys
Pre-Test	1.89	2.18	1.98	2.25
Post-Test	1.67	1.90	1.95	2.44
Change from Pre- to Post-	-0.22	-0.28	-0.03	0.19
	Difference Between Girls' Boys' Mean		Differences Between Girls' Boys' Mean	
Pre-Test	0.29		0.27	
Post-Test	0.23		0.49	

3.3 Attitudes Regarding Self-Efficacy and Prevention

The third section of the *MVP Survey* includes a second unidimensional scale developed to assess student attitudes about their personal ability to prevent and/or intervene in situations involving male violence against women (the “SEV Scale”). It is hypothesized that students with more confidence in their ability to effect change in these situations will be empowered to play an active role in reducing gender violence by being active bystanders and educating their peers. An example of an item in this scale is, “A group of guys would listen to me if I confronted them about their sexist behaviors.” The eleven statements in this section of the survey are followed by a five point, Likert-type scale ranging from “Strongly Disagree” to “Strongly Agree.” **Higher scale scores suggest a greater level of self-efficacy.**

As was the procedure for analyzing the previous unidimensional scale, survey responses in this section were coded, entered into a database, and analyzed using SPSS. Validity and reliability testing was completed. The reliability of the SEV Scale was measured using Chronbach’s alpha. For the pre-test measure, $\alpha = .74$. Chronbach’s alpha for the post-test measure was $\alpha = .84$. Principal Component Analysis was employed to assess construct validity. Factor loadings for the SEV Scale once again suggest that the 11 individual items fit well within the scale. The principal component scale explains close to 30 percent of the variation of the 11 original items for the pre-test measure and 40 percent of the variation for the post-test measure. Because these explained variations are again somewhat average, a scale comprised of the mean values of the 11 survey items was created to ensure that results were robust to changes in scaling procedures.

The dependent measure for the SEV Scale is of an interval nature; therefore, multivariate regression analysis techniques were once again employed to study between-group effects. Two regression models were run using the newly created variable representing the average across the 11 measures. One regression was run without including any independent variables. Another was run controlling for several independent variables, including the Pre SEV mean score, group assignment (treatment vs. comparison), gender, race, age, and survivor status.

As was the case for the AV Scale results described above, strong treatment effects were found for this scale. As was true for Years 1 and 2, the MVP Program achieved its desired goal of improving students’ confidence in their ability to confront violent and sexist behaviors. Specifically, students who received MVP training in Year 3 displayed higher self-efficacy scores than comparison group students, and the difference was statistically significant ($p = .000$). For students in the treatment group, mean scores increased from 3.57 at pre-test to 3.88 at post-test. For comparison group respondents, however, mean scores decreased slightly from 3.44 at pre-test to 3.41 at post-test. Mean scores representing the change over time for both groups are presented in Table 3.3.1.

Table 3.3.1		
SEV SCALE		
Mean Score by Group		
<i>(Higher mean scores indicate higher levels of self-efficacy)</i>		
	Treatment Group Mean	Comparison Group Mean
Pre-Test	3.57	3.44
Post-Test	3.88	3.41
Change from Pre- to Post-Test	0.31	-0.03

Analysis of the Year 3 survey results did not show any statistically significant distinctions between young women and young men in terms of their confidence to be active bystanders. Both boys and girls who received MVP training improve markedly from pre- to post-test. While girls had a slightly higher mean score on self-efficacy at post-test (3.98 for girls compared to 3.77 for boys), the difference was not statistically significant. This result is similar to the Year 1 evaluation results but differs from Year 2 results where girls appeared to have a particularly strong reaction as compared to boys.

For the first time, significant differences were found in Year 3 regarding self-efficacy scores of African American and Latino students. African American students scored significantly lower on the self-efficacy scale than did Caucasian students ($p = .001$) as did Latino students ($p = .012$), indicating that in Year 3 the MVP Program was not as effective at increasing the self-confidence among students in these two racial/ethnic groups. Again, though, this finding is not consistent with findings from the previous years of program evaluation and is difficult to interpret given the differences in the racial/ethnic composition of the treatment and comparison groups. As a result, the finding should not be interpreted as conclusive evidence that MVP is more effective for certain racial/ethnic groups than for others.

3.4 Peer Norms: Student Perceptions of Self and Others

While a majority of questions on the *MVP Survey* were the same for boys and girls, Part IV of the survey was designed to gauge girls' and boys' responses separately. In this section, young women and men are asked to respond to two sets of seven statements. These statements are different in the female and male versions of the survey. Each statement is followed by a five point, Likert-type scale response set ranging from "Strongly Agree" to "Strongly Disagree." The first set of statements (Part A) asks students to describe *their own* attitudes regarding gender violence. The next set of questions (Part B) includes the same seven statements, but students are asked to choose the response that indicates how they think "*the average male [or female] student at your school*" would respond. This year, the statements were analyzed as if they comprised a single scale in order to see if any statistically significant differences emerged. **For these statements, lower mean scores correspond with ignorant or sexist attitudes and are undesirable from a programmatic standpoint.**

This section of the survey was designed in an attempt to understand whether students judge or perceive their peers as being more sexist, or more accepting of sexist ideas and violent behaviors

toward women, than themselves. The hypothesis is that students typically do have this perception of their peers, and that this may contribute to the pressure that many young men and women feel to behave in inappropriate or unhealthy ways. The idea that men do not (or should not) care about women, or that they are not (or should not be) sensitive to the sexist images, stereotypes, and behaviors around them is part of what MVP tries to deconstruct during its training. For example, MVP re-frames the issue of male violence against women as a *man's issue* and emphasizes the potential of males as helpful bystanders rather than potential perpetrators. MVP also provides young people with the opportunity to discuss their feelings about these issues with their peers. In terms of Program impact, one would hypothesize that after MVP training, participants would have more complete knowledge about (and therefore an improved perception regarding) the “real” attitudes of their peers. Again, this survey section was meant to test that hypothesis.

Because there was such a low number of respondents for testing in each group once the sample was split between treatment and comparison groups and further divided by gender, the power of the statistical analysis of the data is low. In other words, it is likely that statistically significant results could not be detected with such small samples sizes. Nonetheless, regression analyses were run for each of the four sets of scales in this survey section (i.e., female self-perception, male self-perception, female perception of others, and male perception of others). The findings are presented in the following paragraphs.

Boys' Perceptions of Self and Others

At pre-test, the mean score for treatment group males was higher (3.41) than that of the male comparison group (3.34); however, this difference was not statistically significant ($p = .574$). At post-test, the mean score for boys who received MVP training increased (3.74) and remained higher than that of their peers in the comparison group whose mean score decreased at second testing (3.20). This difference between groups at post-test was significant ($p = .000$). This finding corresponds with and reinforces the AV Scale findings presented in an earlier section of this report regarding the effectiveness of MVP at improving the attitudes of students who receive training.

However, as mentioned earlier, the purpose of this section on the survey was to understand whether MVP effectively changes students' perceptions of their peers. At pre-test, boys in both the treatment and comparison groups perceived their male peers to hold more sexist attitudes than themselves (see Table 3.4), as was expected. Boys in the MVP group scored a 2.43 on this scale, while boys in the comparison group had a mean score of 2.95 at pre-test, and the difference was significant ($p = .000$). This means that, prior to training, the boys who received the MVP Program in Year 3 perceived their male peers as being more sexist than did the boys in the comparison group. At post-test, the boys who received MVP training did not perceive their peers as being any less sexist than they did at pre-test. The mean score for boys who did not receive MVP training declined slightly at pre-test, but the difference was not significant. Moreover, the difference between boys in the treatment group and boys in the comparison group at post-test remained significant ($p = .004$), meaning that boys who received MVP training continued to view their male peers as more sexist than did males in the comparison group. In other words, after participating in the MVP Program, treatment group males did not have more positive perceptions of the attitudes and beliefs held by other males in their peer group. This finding appears somewhat contrary to what was observed in Year 2. However, as explained, those descriptive results were not presented as being statistically significant or conclusive.

Table 3.4				
Survey Section IV				
Mean Scores by Gender and Group				
(Lower mean scores indicate ignorant or sexist attitudes about gender violence)				
	Pre-test Own	Post-test Own	Pre-test Other	Post-test Other
BOYS				
Treatment Group	3.41	3.74	2.43	2.45
Comparison Group Mean	3.34	3.20	2.95	2.82
Significance	p = .574	p = .000	p = .000	p = .004
GIRLS				
Treatment Group Mean	3.81	4.05	3.02	3.10
Comparison Group Mean	3.58	3.73	3.18	3.26
Significance	p = .017	p = .003	p = .193	p = .220

Girls' Perceptions of Self and Others

At pre-test, the mean score for treatment group girls was higher (3.81) than that of the girls in the comparison group (3.58). This difference was significant at pre-test ($p = .017$), meaning that girls in these two groups were not similar on this scale from the beginning in that girls in the MVP group were less accepting of sexist behaviors. At post-test, the mean score for girls who received MVP training increased to 4.05 and remained higher than that of their peers in the comparison group whose mean score increased at second testing to 3.73. Again, the difference between the two groups in terms of their own attitudes was significant ($p = .003$). As was true for boys, this result confirms that the MVP Program was particularly effective in improving the attitudes of students who received training.

At pre-test, girls in both the treatment and comparison groups perceived their female peers to be more accepting of sexist attitudes than themselves (see Table 3.4). Girls in the MVP group had an average score of 3.02 on this scale, while girls in the comparison group had a mean score of 3.18 at pre-test. This difference between treatment and comparison groups was not significant ($p = .193$). At post-test, both the girls who received MVP training and those who did not perceived their peers as being only slightly less accepting of inappropriate or sexist behaviors than they did at pre-test. The difference between the two groups at post-test was not significant ($p = .220$). According to Year 3 survey results, receiving MVP training was not associated with improved perceptions of peer attitudes and behaviors for girls.

Year 3 results from this section of the survey did not seem to support the hypothesis that MVP training helps to change student perceptions about peer group attitudes and norms. This is different from descriptive data from the Year 2 survey, which suggested that MVP may be effective in this manner. However, those preliminary results from Year 2 were not statistically significant. It may be the case that girls and boys who receive MVP training have a heightened awareness of and sensitivity to gender bias and sexist behaviors, and they may feel that peers who have not benefited from such training must still hold ignorant attitudes. Or perhaps MVP could strengthen its programming in some way to debunk the idea that most people are personally comfortable with sexist attitudes and behaviors. It may also be the case that the scale on the *MVP Survey* that is used to measure this construct needs to be improved. Given the small sample sizes and mixed results across two testing periods, it is not prudent to make conclusive statements about the effectiveness of MVP programming on this construct.

IV. STUDENT SATISFACTION

V. SUMMARY OF FINDINGS ACROSS EVALUATION STUDIES

As mentioned, the multi-year, multi-method approach to understanding the outcomes and effectiveness of the MVP Program sets it apart from the typical evaluation studies of gender violence prevention programs reported in the literature. For example, few published studies of teen dating violence prevention programs feature student voices in terms of explaining how gender bias and violence affect their daily lives or describing their reactions to prevention programming. To date, extensive qualitative case study data have been collected and reported that detail the need for and impact of the MVP Program in high schools. These data include numerous in-depth interviews with MVP student participants, multiple observations of program training, and several key informant interviews. Additional qualitative data collected as part of the evaluation of MVP include focus groups with female participants as well as three years of participant satisfaction data. Analysis of these qualitative data have helped to contextualize and better interpret the three years of survey data that have also been collected as part of the evaluation of the MVP Program.

The *MVP Survey* was created specifically for the MVP Program and, over the course of three years of administration, has been found to be a valid and reliable measure of student knowledge, attitude, and self-efficacy. In Year 1, the *MVP Survey* was administered as a pre- and post-test assessment of MVP participants. Findings from the analysis of the first year of survey data were positive, showing MVP to have effectively reached its objectives of improving student knowledge, awareness, and attitudes. However, the Year 1 study design did not incorporate a comparison group. For the Year 2 study, a comparison group was drawn using a number of schools that wanted MVP training for their students but were on a wait list. Again, evaluation findings using this stronger design suggested that students who received MVP training experienced increased levels of knowledge and awareness, improved attitudes regarding violence against women, and increased confidence in their ability to prevent and/or intervene in abusive or violent situations. As a result of using a quasi-experimental design in the Year 2 study, there was increased confidence that observed changes were the result of MVP training and not some other influence. In Year 3, a quasi-experimental pre- and post-test design using comparison group was again employed, and study findings were replicated. For the third year in a row, results indicated statistically significant improvements in knowledge, attitude, and self-efficacy for MVP students at post-test as compared to students from a comparison sample.

An important focus of the multi-year, multi-method evaluation of the MVP Program was to analyze results in terms of any gender differences that might emerge. In particular, Program administrators were interested in understanding whether participation in the Program and the Program's emphasis on bystander education might disempower or be inappropriate or uncomfortable for female students. Somewhat mixed results merged across the three years of evaluation. In terms of attitude change, young women enter and exit the MVP Program training with better (less sexist or less accepting of sexist and violent behaviors) attitudes than young men and, based on mean scores, their improvement from pre-to post-test is less marked than that of male participants. This finding was consistent across studies. In terms of self-efficacy or confidence to prevent or intervene in abusive or violent situations, young women exhibited a higher level of self-efficacy than young men at both pre-and post-test in Years 2 and 3. This is a good indication that MVP's programmatic approach is helpful to and effective

for female participants. Though these differences were not confirmed as statistically significant, focus group data from the Year 2 evaluation indicated that MVP's bystander approach is perhaps exactly what is responsible for the feelings of safety and empowerment that young women feel after training.

Overall, findings from the third year study of the MVP Program confirm what has been tested in previous phases of the evaluation. That is, students who are exposed to the MVP curriculum and training are more highly aware of and knowledgeable about gender violence, have less sexist or inappropriate attitudes concerning violence and harassment against women, and are more comfortable and confident in their ability to intervene in situations involving gender harassment and violence. Moreover, students continue to report being highly satisfied with their experience in the MVP Program, including the format and content of the MVP training. These outcomes, consistent across all three years of the evaluation, provide strong evidence that MVP is effective in meeting its objectives.

This approach to evaluating the MVP Program, though very strong, could be improved. Specifically, the use of larger comparison samples drawn from a randomized control group design would improve our ability to discern and interpret statistically significant gender and racial/ethnic differences in terms of program impact. As well, MVP Program staff and funders may be interested to know more about how well MVP students retain the positive effects after training has ended. A longitudinal study of MVP would be valuable in that regard. Nonetheless, given limited resources, MVP has done an effective and systematic job documenting quantitative and qualitative outcomes of their efforts across multiple years and can be confident that the positive outcomes shown can be attributed to the MVP curriculum and training.