Findings from an Innovative Teen Pregnancy Prevention Program

Evaluation of The Web of Life in New Mexico

Final Impact Report for

National Indian Youth Leadership Project

November 30, 2015

Prepared by

Susan Lee Carter, Susan Lee Carter Inc.

Jan Vanslyke, Evaluation Specialists

Blair Beadnell, Evaluation Specialists

Carter SL, Beadnell B, Vanslyke J. Evaluation of the Web of Life teen pregnancy prevention program: Findings from an innovative positive youth development approach for American Indian youth. National Indian Youth Leadership Project. 2015.

Acknowledgements:

Many people made this evaluation possible. First, thanks go to the National Indian Youth Leadership Project and its founder and Executive Director, McClellan Hall, for his unswerving support of Native youth and of high quality program evaluation. Thanks go to Susan Carter as principal investigator of the Web of Life study and for her twenty years of service and support. The Web of Life program was professionally and lovingly implemented and supported by the very talented staff at the National Indian Youth Leadership Project: Matthew Baker, Natasha Clark-Baker, Neal Ferris, Shurvina Lee, Devin Seowtewa, Nat Stone, Gloria Tom, and Heather Yazzie Campbell. We are also very grateful to the staff and students at participating program intervention and comparison sites New Mexico. Their dedication to positive youth development was critical to the success of the study.

The Web of Life Project was reviewed and approved by the Navajo Nation Human Research Review Board, Protocol #NNR 11.320
Beverly Becenti-Pigman, Chair
Navajo Division of Health
PO Box 1390
Window Rock, AZ 86515
928-729-4415

This publication was prepared under Grant Number TPA2AH000002 from the Office of Adolescent Health, U. S. Department of Health & Human Services (HHS). The views expressed in this report are those of the authors and do not necessarily represent the policies of HHS or the Office of Adolescent Health.

EVALUATION OF THE WEB OF LIFE IN NEW MEXICO: FINDINGS FROM AN INNOVATIVE TEEN PREGNANCY PREVENTION PROGRAM

I. Introduction

A. Introduction and study overview

Disparities in health outcomes have been widely acknowledged among American Indian communities nationally and, specifically, within the two target communities examined in this report. In its 2009 Racial and Ethnic Health Disparities Report Card, the New Mexico Department of Health identified a number of health disparities for American Indians in New Mexico and the level of intervention required to address them. Disparities requiring urgent interventions included births to American Indian teens ages 15 to 17. After declining for more than a decade, the teen birth rate among American Indians increased 12% between 2005 and 2007, which was more than any other racial/ethnic group.[i] Although birth to all teens in New Mexico fell 35% over the past decade, it still has the second highest rate in the U.S., especially in rural counties, where most American Indian teens reside.[ii]

Prevention during the early teen years is needed, yet, only a limited number of programs are available that have been designed specifically for Native teens, and none have been rigorously evaluated. There is also limited research on the best ways to reduce teen pregnancy in the Native population. The Web of Life (WOL) program, which adapts Project Venture to teen pregnancy and sexually transmitted infection prevention, is designed to address these gaps.

The WOL study was designed in part to support the Office of Adolescent Health's (OAH) Teen Pregnancy Prevention initiative's goal to support innovative pregnancy prevention strategies for youth that are both medically accurate and age appropriate. The Funding Opportunity Announcement (FOA # OPHS/OAHTPP PREP Tier2-2010) stated that the Tier 2 initiative was designed specifically to support research and demonstration programs that would, "develop, replicate, refine, and test additional models and innovative strategies for preventing teenage pregnancy." The WOL program is significantly adapted from the evidence-based (SAMHSA NREPP,[iii] DOJ Crime Solutions[iv]) Project Venture, which has demonstrated success in preventing substance abuse, HIV, suicide and related problems among adolescent

Native youth, as well as in promoting resilience and related positive youth development outcomes. The developers believed that the WOL adaptation to Project Venture held great promise as a teen pregnancy prevention approach and welcomed the opportunity to further test and refine this innovative approach, which includes adventure based experiential activities carefully sequenced throughout the year in order to strengthen life skills and resilience factors of participants. Major adaptations included a greater focus on healthy relationships and an equine component, known as Native American Horse Inspired Growth and Healing.

After a scan of needs, gaps and offerings in services related to teen pregnancy prevention in the communities within its service area, the National Indian Youth Leadership Project (NIYLP) approached two Native communities in west central New Mexico and negotiated agreements to implement WOL among 6th-grade youth (11-12 year olds) in two target schools. Sixth grade was chosen because of the potential to impact youth before they begin to be engaged in sexual behavior. This is in keeping with the guiding principles of NIYLP and Project Venture of engaging with youth to build resilience before troubling behaviors develop. This report will describe the implementation and impact of WOL.

B. Primary research question(s)

The primary research question asks,

> Immediately at the end of the treatment, what is the impact of WOL relative to a no-treatment matched comparison group on having ever had sex?

C. Secondary research question(s)

The secondary research question examined the impact of WOL six months after the end of the treatment. Cohorts 1 and 2 were included in this analysis (Cohort 3 was not included in this analysis since that data collection period fell after the end of the grant period). The secondary research question asks,

> Six months after the end of the treatment, what is the impact of WOL relative to a no-treatment matched comparison group on having ever had sex?

The primary and secondary research questions were also examined to determine whether the impact of WOL varied by the sex of participants.

II. Program and comparison programming

A. Description of program as intended

WOL is adapted from NIYLP's Project Venture, an evidence-based program designed to prevent substance abuse and promote resilience and mental health among American Indian adolescents. Project Venture is a culturally guided, adventure-based experiential youth development program developed over 30 years of work with high-risk American Indian youth. The model relies on American Indian traditional values, customized for each target community, to help youth develop positive traits including belonging, mastery, independence, and generosity.[v] These traditional values align well with mainstream notions of positive youth development including competence, confidence, connection, character, contribution and caring.[vi]

WOL shares Project Venture's guiding principles which provide adventure-based group experiential activities that are carefully sequenced over a year's time and are guided by local cultural values. For example, giving back to one's community is an important value and is implemented through community-based service learning projects. During the planning phase of this study, WOL activities and reflection questions were designed to focus on healthy relationships, respect for self and others, and healthy decision making. WOL includes an equine component known as Native American Horse Inspired Growth and Healing to further these goals. WOL serves youth in the 6th grade, primarily 11- and 12-year-olds, consistent with the primary target age for Project Venture. Because our research informs us that youth in this age range in our communities are not engaging in sexual intercourse to any measurable degree (between 1 and 2%), WOL does not provide direct instruction regarding sexual and reproductive health. Rather the focus is on building upon the strengths of participants in preparation for later challenges.

Underlying program theory. WOL is based on a positive youth development framework which views young people from a strengths-based approach, in contrast to prevalent deficit-based models. WOL seeks to identify, strengthen, and support positive features of young people through culturally meaningful approaches which are adventure-based, experiential, and which include a reverence for the natural world

and a community-based service learning ethic. The emphasis is on developing self-awareness and healthy relationships in order to prevent a host of potential adolescent problems such as teen pregnancy, substance use, depression and anxiety. As positive aspects of youth are developed, it is hypothesized, youth become more capable of healthy decision-making and display fewer risky social and sexual behaviors, as well as improved/positive group dynamics with an emphasis on cooperation and mutual respect within and across genders.

Intended program content. WOL content is presented in the curriculum manual which includes descriptions for each activity. Reflection questions are included at the end of each session to help participants extend their understanding of how to apply lessons learned in each session. The curriculum guide also includes approximate time, materials, and other resources required for each session.

Intended program components. WOL includes a number of components delivered by specially trained NIYLP staff in school, community, and wilderness settings as follows:

Classroom-based component. Sessions are delivered experientially/interactively and are sequenced developmentally throughout the school year, increasing in challenge and skill level. All participants are enrolled in these sessions, which are designed to have positive impacts on their own, as well as serve as preparation for out of school time activities for those participants who choose those components.

Weekly after-school component. Youth participating in the classroom based component can also elect to participate in the out of school time components, including the weekly after-school sessions which take place after the school day. These sessions build on the classroom based sessions and provide further preparation for the more challenging daylong and multiday components. These are most often located at the school campus or nearby locations.

Weekend component. Youth may also participate in daylong sessions on weekends and school holidays. These sessions provide additional skill building challenges and include activities such as hiking, rock climbing, rappelling, caving, and canoeing. These are most often located at off-site locations accessible within one day.

Multiday component. Youth may also participate in this component, which includes the highest level of challenge and skill for participants. One of these sessions is Native American Horse Inspired Growth and Healing during which participants interact with horses in order to deepen their understanding of healthy relationships, communication skills, and interpersonal awareness in a culturally meaningful setting. Other multiday events include canoeing or rafting on rivers in the region, such as northern New Mexico and the Grand Canyon in Arizona. Activities can be modified depending on location, weather, and resources for future adoption sites.

Intended program dosage. WOL delivers services to 6th-graders in selected program sites scheduled regularly throughout the school year, which typically runs from mid- to late-August through May. Table II.1 illustrates all program components with the ideal range of dosage for each. School-based component dosage may vary across sites depending on class duration at each site. Out of school time dosage varies depending on factors such as the type of activity, travel distance to activity settings, and length of holiday weekends.

Table II.1. Web of Life program components

Component	Number of Sessions	Intended Duration of Each Session	Minimum Annual Hours per Participant	Maximum Annual Hours per Participant
Classroom/School	26	.75-1.0 hours	19	26
After-School	20	1.5-2.0 hours	30	40
Weekends	10	6-8 hours	60	80
Native American Horse Inspired Growth and Healing	4	6-8 hours	24	32
Multiday Trip 1	1	3-4 days *	48	64
Multiday Trip 2	1	5-6 days *	80	96
Total			261	338

* Days calculated at 16 hours

B. Description of counterfactual condition

The study's counterfactual condition was considered a "no treatment/ business as usual" design. Comparison schools reported little or no services related to teen pregnancy prevention in the 6th or 7th grades; however, one of the comparison schools provided additional programs and services after the study began such as yoga instruction, cooking and nutrition activities provided by an outside agency, and an

intensive focus on ethnic identity. While none of these would be considered "adventure-based" activities, they share with WOL the fact that they are experiential and culturally focused.

III. Study design

A. Sample recruitment

Setting and context. The study was implemented by the NIYLP in five middle schools and their communities located in several counties in west central New Mexico with high American Indian populations. NIYLP had a successful history working in the two intervention communities and one of the comparison communities (two of the comparison communities were relatively new to NIYLP). During the first two years of the five year grant, program and evaluation teams piloted every aspect of the study including identification of appropriate study sites, instrument development, and curriculum adaptations. The formal study took place in three cohorts during years three through five (fall 2012 through spring 2015).

Target population and recruitment process. The target population for the study was American Indian 6th grade youth enrolled in middle schools throughout NIYLP's local service area. NIYLP staff held meetings with potential school districts to determine suitability, interest, and access. All 6th grade youth enrolled in each of the five study schools at baseline (beginning of 6th grade school year) were eligible to participate if they were capable of completing the study survey instrument and if their parents/guardians consented to their participation. In the first year of the study, each student who returned a consent form received a $10 incentive. In subsequent years, cash incentives of $10 per student were paid to teachers to encourage the return of consent forms.

B. Study design

The two program intervention sites were identified based on their commitment and capacity to support the full implementation of WOL. NIYLP had previously implemented Project Venture in both sites and was confident in the ability of these two sites to successfully support the intervention.

At least one comparison school site was selected for each program intervention school site. Comparison schools were selected to maximize their similarity in terms of type and size of school, tribal

affiliation, and rurality. As noted above, one of the comparison schools appeared to provide a significant amount of enrichment services, so a sensitivity analysis was conducted removing that site from the analysis.

C. Data collection

1. Impact evaluation

Participants completed a paper and pencil survey developed and piloted during the start-up phase of the study. The survey was administered at baseline immediately before the intervention began, posttest at program exit, and at a 6-month followup (Cohorts 1 and 2 only) mid-way through the 7th grade year. The followup data collection point for Cohort 3 fell outside of the grant period and was not included in the analyses. Survey administration took approximately 30 minutes. Appendix A displays information about impact data collection efforts and timing.

At each data collection point, the survey was read aloud to participants in classroom settings by a survey administrator trained by the program evaluator and who was not a direct program facilitator. One make-up session was scheduled for each school if there were absences on the day of the survey. A tear-off cover sheet was used to ensure confidentiality. Classroom teachers provided incentives to keep the youth focused on survey completion, such as free time after the survey. The timing and mode of survey administration was the same for both program and counterfactual conditions.

2. Implementation evaluation

The implementation evaluation was designed to provide context for the impact findings by assessing adherence to and quality of the program model, the experiences of the comparison participants, and overall contextual conditions. Appendix B contains additional information about the implementation evaluation plan.

Adherence. Program facilitators completed a fidelity log after each session. These fidelity logs listed every activity for every session. Facilitators documented whether activities were fully, partially, or not implemented, and if any changes were made to the session plan. All changes to the original plan were noted for each session. Each fidelity log was accompanied by an attendance sheet.

Annual performance reviews were conducted by the WOL program coordinator or the NIYLP director of operations to assess strengths and challenges of each employee, and to insure that all required trainings and certifications had been completed.

Quality. Quality of implementation was assessed by means of observations of a random selection of 10% of all sessions, stratified by session type to assure that all four types of activities were adequately assessed-school-based, after-school, weekend, and multiday sessions. Additional observations, beyond the required 10%, were conducted by the evaluation team.

Annual midyear interviews with stakeholders provided another perspective on implementation quality. Stakeholders included youth participants, parents, teachers, and school administrators. In individual interviews, they were asked questions related to how much they knew about the program, what changes they had observed in students or in themselves as a result of participation in WOL, and if they had suggestions for improvement. All observations and interviews were conducted by a member of the evaluation team designated as the independent observer using the same protocol each year.

Counterfactual experiences. The survey administrator and evaluation director had frequent contact with **comparison** schools especially during survey administration time periods and noted important factors which could impact outcomes such as whether other sexual and reproductive health programs were being offered or the presence of positive youth development/enrichment activities. In most cases, staff did not have the ability to change these experiences. However, in one case, we requested that a school change their implementation of a well-known teen pregnancy prevention program to serve older youth at another school to reduce contamination. They complied with this request. This occurred before programming began at the beginning of the study.

Context. Contextual factors such as changes in school schedules, important events in the community, severe **weather**, and changes in key personnel at school sites were identified and discussed at weekly project staff meetings especially when seen as potentially impacting program implementation or outcomes. Data came from school calendars, websites, and newsletters; and from local news reporting, and staff observations and conversations with youth and adult community members.

D. Outcomes for impact analyses

The primary and secondary outcomes were having ever had sexual intercourse, measured at posttest and six-month followup, respectively. Tables III.1 and III.2 describe how the primary outcomes were operationalized using survey data. Given the young age of the participants, we expected that the percentage having sexual intercourse would be very low. Because of this, we also examined outcome variables we presumed could be affected by the treatment and that are found in research to be factors associated with eventual sexual behavior. Tables III.3 through III.6 describe how the other research questions were operationalized.

Table III.1. Behavioral outcome used for primary impact analysis research question

Outcome name	Description of outcome	Timing of measure relative to program
Having sexual intercourse	The variable is a yes/no measure of whether a person has ever had sexual intercourse. The measure is taken directly from the following item on the survey: • "Have you ever had sexual intercourse?" The variable is constructed as a dummy variable where respondents who respond yes they have had sex are coded as 1 and all others are coded as 0.	Immediately after program ends

Table III.2. Behavioral outcome used for secondary impact analysis research question

Outcome name	Description of outcome	Timing of measure relative to program
Having sexual intercourse	The variable is a yes/no measure of whether a person has ever had sexual intercourse. The measure is taken directly from the following item on the survey: • "Have you ever had sexual intercourse?" The variable is constructed as a dummy variable where respondents who respond yes they have had sex are coded as 1 and all others are coded as 0.	6 months after program ends

Table III.3. Other outcomes used to measure changes in attitudes and behaviors that are associated with sexual behaviors

Outcome name	Description of outcome	Timing of measure relative to program
Positive youth development	The variable consists of 4 subscales and a global measure of Internal Assets, created by the Search Institute (Developmental Assets Profile).[vii] Each of the four subscales consists of 32 individual statements as follows: • Commitment to Learning (7 statements) • Positive Values (11 statements) • Social Competence (8 statements) • Positive Identity (6 statements) Respondents are asked to check if the statement is true now or within the past 3 months: • 0=not at all or rarely • 1=somewhat or sometimes • 2=very or often • 3=extremely or almost always	Immediately after program ends 6 months after program ends
Mental health	The Mental Health measure is composed of 2 sub-measures adapted from the Behavior Symptoms Checklist (Derogatis et al.)[viii] • depression/suicidal ideation (7 statements) • anxiety/fearfulness (7 statements) Each statement is scored from 0 to 3 indicating how true the statement is for the respondent during the past week: • 0=not at all true • 1=a little true • 2=pretty much true • 3=very much true	Immediately after program ends 6 months after program ends
Cultural factor	Ethnic Identity is assessed using the Multigroup Ethnic Identity Measure (MEIM-R) (Phinney, J.).[ix] It consists of 6 statements addressing exploration of and commitment to ethnic identity. Each statement is scored from 0 to 4 indicating how much respondents agree or disagree with the statement: • 0=strongly disagree • 1=disagree • 2=neutral • 3=agree • 4=strongly agree	Immediately after program ends 6 months after program ends
Substance use	The variable is a yes/no measure of whether the respondent has ever used cigarettes, alcohol (including binge drinking), and marijuana. The measure is taken directly from the following four items on the survey: During the past 30 days, have you… • "smoked a cigarette, more than just one puff?" • "had a drink of alcohol, more than just a sip?" • "had five or more drinks in a short time?" • "smoked marijuana?" The variable is constructed as a dummy variable where respondents who respond yes, they have used, are coded as 1 and all others are coded as 0.	Immediately after program ends 6 months after program ends

E. Study sample

Five schools in west central New Mexico made up the study sample. Two schools received the WOL intervention, while three were business-as-usual comparisons. In each school a majority of the enrolled students were American Indian. All participants were in the 6th grade at the start of the study.

The total number of youth eligible for participation in the study was 1,089 (399 treatment and 690 comparison). Consent rates varied widely across sites, but once consent was granted, study retention rates were relatively high and constant through the followup in grade seven. Consent was obtained from 783 youth (250 treatment and 533 comparison) resulting in a 72% overall consent rate (63% treatment, 77% comparison).

A total 1,089 youth were eligible to participate in the study in the baseline and posttest surveys and 746 youth were eligible to participate in the followup survey (Cohort 3 did not participate in the followup survey). Of eligible youth, 738 (68%) completed the baseline survey, 619 (57%) completed the posttest survey, and 359 (48%) completed the followup survey. These response rates are based on the total number of eligible youth in the five schools before consent was obtained. Appendix C displays the flow of study participants throughout the three years of the study.

F. Baseline equivalence

We selected the individuals with complete baseline and posttest data (n = 598; 21 cases had missing demographic or outcome data), and examined baseline equivalence in the immediate posttest and follow up analytic samples. To do this, we ran a separate analysis for each participant characteristic in which it was the outcome, and experimental condition was the predictor. We used Generalized Estimating Equation analyses with robust estimators to adjust findings for the clustering inherent in this study design. The analyses used a linear model for continuously measured characteristics and a linear probability model for dichotomously measured characteristics. For race/ethnicity, we created two variables (self-identification as American Indian and as Hispanic) to represent the two key ways race/ethnicity varied in this sample. We found baseline nonequivalence ($p < .05$) on some demographic and baseline measures of outcomes.

Because of the nonequivalence between conditions, we performed a series of Exact Matching and Propensity Score Matching procedures to create more equivalent analyses samples. We performed the matching separately for each cohort, and used the R MatchIt utility. Following the match, we merged the matched cohort datasets. All matching models allowed greater than one comparison participant to be matched to each treatment participant.

Appendix E provides details about the matching models and their relative success. In the full sample of five schools, only one model achieved equivalence across all demographics and baseline-measured outcomes (see Table III.4). The resulting matched dataset served as our benchmark in all analyses. Because our impact analyses included separate analyses for females and males, we examined equivalence within sex. Findings showed baseline equivalence on all demographic characteristics and baseline-measured outcomes for females and males.

Table III.4. Summary statistics of key baseline measures for youth in matched benchmark sample

Baseline measure	Treatment mean or % (standard deviation)	Comparison mean or % (standard deviation)	Treatment versus comparison mean difference	Treatment versus comparison *p*-value of difference
Cohorts 1, 2 and 3 (For use in posttest analysis)				
Behavioral measure				
Ever had sex	<1%	<1%	<1%	.73
Other measures				
Gender (female)	53%	57%	-4%	.49
Race/ethnicity: American Indian	87%	88%	-1%	.78
Race/ethnicity: Hispanic	15%	17%	-2%	.70
School type: Pueblo (Navajo)	30%	35%	5%	.24
Age	11.75 (0.60)	11.76 (0.58	-.01	.88
Used substances[a]	6%	8%	-2%	.55
Internal assets: Learning	2.42 (0.49)	2.36 (0.57)	.06	.27
Internal assets: Values	2.36 (0.46)	2.28 (0.55)	.08	.10
Internal assets: Social	2.20 (0.54)	2.26 (0.54)	-.06	.10
Internal assets: Identity	2.21 (0.55)	2.28 (0.55)	-.07	.16
Internal assets: Total Score	2.30 (0.43)	2.29 (0.49)	.01	.78
Cultural factor: Ethnic Identity	2.80 (0.84)	2.76 (0.95)	.04	.67

Baseline measure	Treatment mean or % (standard deviation)	Comparison mean or % (standard deviation)	Treatment versus comparison mean difference	Treatment versus comparison *p*-value of difference
Mental health measure: Anxiety	1.01 (0.64)	0.90 (0.70)	.11	.13
Mental health measure: Depression	0.72 (0.60)	0.66 (0.66)	.06	.30
Sample size	**215**	**276**		
Cohorts 1 and 2 (For use in followup analysis)				
Behavioral measure				
Ever had sex	<1%	<1%	<1%	.81
Other measures				
Gender (female)	48%	54%	-6%	.26
Race/ethnicity: American Indian	84%	87%	-3%	.59
Race/ethnicity: Hispanic	14%	16%	-2%	.63
School type: Pueblo (Navajo)	30%	38%	-8%	.12
Age	11.80 (0.63)	11.78 (0.59)	.02	.81
Used substances[a]	6%	8%	-2%	.56
Internal assets: Learning	2.35 (0.51)	2.34 (0.60)	.01	.73
Internal assets: Values	2.32 (0.49)	2.25 (0.58)	.07	.21
Internal assets: Social	2.18 (0.55)	2.25 (0.57)	-.07	.30
Internal assets: Identity	2.20 (0.55)	2.28 (0.58)	-.08	.21
Internal assets: Total Score	2.27 (0.45)	2.27 (0.52)	0	.78
Cultural factor: Ethnic Identity	2.80 (0.82)	2.77 (0.99)	.03	.88
Mental health measure: Anxiety	0.95 (0.62)	0.84 (0.68)	.11	.13
Mental health measure: Depression	0.69 (0.60)	0.63 (0.63	.06	.40
Sample size	**146**	**207**		

[a]Used alcohol, marijuana, or tobacco

G. Methods

1. Impact evaluation

Analytic samples. The immediate pretest analytic sample included data from all three cohorts in school years 2012–2013, 2013–2014, and 2014–2015 and the followup analytic sample included data from the first two cohorts only. To be included in the analysis, participants needed to have provided baseline and outcome data, and have no missing data on baseline-measured demographic or outcome

variables. See Missing Data section below for information on the number of enrolled participants meeting these criteria. We used IBM SPSS v.22 for all statistical analyses.

Model specification. After performing the matching procedure to achieve baseline equivalence, we conducted outcome analyses separately for the posttest (Cohorts 1, 2, and 3) and followup (Cohorts 1 and 2). We used Generalized Estimating Equation analyses with robust estimators. We specified schools within each cohort as subject (i.e., cluster) variables to account for the nested nature of this study's design. In other words, we treated the data as having 15 clusters (the 5 sites X the 3 cohorts). We used the linear model for continuously measured outcomes, and the linear probability model for dichotomously coded (0/1) outcomes. For the analysis of each outcome, we included as predictors the baseline measure of that outcome, condition (treatment = 1, comparison = 0), and demographics (being female, being American Indian, being Hispanic, attending a Pueblo school, and age). In terms of the sexual intercourse outcome, while this was measured at all timepoints as "having ever had sexual intercourse," by controlling for answers at baseline, the probability test would reflect the initiation of intercourse. Any differences between the treatment and comparison groups on the outcomes were considered statistically significant if the *p*-value was less than 0.05, using a two-tailed test.

We also examined whether treatment vs. comparison group effects differed depending on sex of participant (female vs. male). We did this by adding to the above analyses an interaction term predictor, which was the interaction of experimental condition and sex. If the interaction term was significant, we then performed separate Generalized Estimating Equation models for females and males to better understand its meaning.

Missing data approach. No cluster-level attrition occurred (i.e., no school ceased participation). At the participant level, only 21 of the 619 participants with pretest and posttest data (the inclusion criteria for being included for the matching procedure) had missing data on a demographic or the outcome variables (as measured at baseline) used in the matching procedure and analyses. Because of the small number of cases with missing data, we excluded them from further analyses, bringing the sample available for the matching procedures to 598 (76% of the enrolled sample).

Sensitivity analyses. The robustness of our analyses of the benchmark sample was examined by conducting the same analyses with two other samples created during our matching analyses. One was a matching performed on a four-school sample (n = 492) that excluded the one comparison school discussed earlier that had active intervention activities occurring. Matching of this sample achieved equivalence overall and within gender subgroups on demographic variables, and all baseline-measured outcome variables except anxiety and depression. As an additional sensitivity sample, we used a dataset created from the full sample that achieved equivalence on most variables, with the exception of being in a Navajo versus Pueblo school, and baseline-measured positive values, anxiety, and depression.

2. Implementation evaluation

Evaluation of the WOL program implementation provided important information regarding adherence to the program model (fidelity), quality of implementation, experiences of the comparison group, and overall study context. Especially during the pilot-test phase, this information enabled staff to improve and fine tune their efforts and provided insights into the mechanisms behind outcome results. Appendix D includes detailed information about the methods used to analyze various aspects of implementation.

Adherence to program model. The total number of sessions was calculated and compared to the intended number to ensure that the number of sessions was being delivered as planned. The average number of sessions attended by participants was also compared to the intended dosage. The number of session objectives delivered was compared to the total number intended, and a percentage was calculated. Staff compliance with training and certification requirements was determined by ratings of supervisors on the NIYLP performance review system. All trainings and certifications were noted in each staff member's file and reviewed for completeness by the supervisor who discussed issues with staff and developed corrective action plans when needed.

Quality of implementation**.** The quality of staff interactions with participants was calculated as the percentage of observed sessions receiving a "high quality" rating by the same independent observer throughout the study. The quality of participants' engagement with the activities presented was calculated

at the percentage of sessions rated as "moderately engaged" or higher by the independent observer using the session observation form. Stakeholder interviews with teachers, principals, parents, and youth participants were analyzed for positive and negative comments, and for suggestions for improvement. Stakeholder feedback was summarized and published in program newsletters and other outreach activities.

Experience of comparison group and overall context. Experiences of both the counterfactual condition as well as of the treatment condition and overall study context were documented through interviews with teachers, principals, and counselors in each study site. Experiences such as new programs or curricula were documented and analyzed as to whether they were a potential threat to the study. Weekly staff meeting notes included any new events or changes affecting the five study schools and communities. Staff determined whether these were minor or significant events.

IV. Study findings

A. Implementation study findings

The implementation study focused on four areas. In general, the analysis found that the WOL program was implemented with high fidelity, quality, and participant engagement.

Adherence to program model. A total of 2,639 activities within 691 program sessions were offered over the three-year study period with an overall completion rate of 89%. Completion rates increased from 84% in Year 1, to 87% in Year 2, to 97% in Year 3. This may have been because the number of activities per session decreased slightly since facilitators felt that they had been a bit unrealistic in the original curriculum design, including more activities per session than could reasonably be executed. While not compromising the overall theme and objectives of each session, staff were able to reduce the number of activities per session which increased the quality of implementation and allowed more time for critical reflection activities at the end of each session. An independent observer attended 72 sessions over the study period, reporting 84% compliance regarding session completion.

Over the three cohorts of the study, 430 youth were enrolled. On average, 88% of students attended the school-based component. Forty-four percent elected to attend both the weekly after-school

component and the weekend component. Seventy-six percent participated in at least one of the three multiday events.

Quality of implementation. Quality of implementation was assessed by ratings provided by the independent observer as well as through stakeholder feedback interviews conducted annually. Observer ratings averaged 4.9 out of 5.0, where the higher the score the higher the quality. The lowest rating (4.5) was assigned to the item related to how well facilitators kept track of time. This information was used to help facilitators ensure that they did not run out of time at the end of each session for critical debriefing and reflection. Stakeholder feedback from participants, parents, teachers, and school administrators was consistently positive throughout the study period.

Three of the four WOL staff facilitators were with the program since its inception, and with the agency for a number of years before that. One of the original staff facilitators left in Year 2 and was replaced by another very experienced NIYLP staff member. Staff performance reviews were generally favorable and staff were able to keep up with required trainings and certifications such as Wilderness First Aid, Wilderness First Responder, and Mental Health First Aid. Every staff facilitator had the opportunity to attend OAH regional or national meetings and trainings at least once.

Experience of comparison group. While none of the three comparison schools offered a formal teen pregnancy prevention program, each briefly addressed sexual and reproductive health through two or three sessions in health, physical education, and wellness classes during the study period when participants were in 6th and 7th grades. One comparison school had a small/low dosage experiential education program for students. As noted in Section II above, another comparison school implemented several enrichment programs and activities for students along with a strong focus on American Indian ethnic identity. Some of these activities were similar to those offered through WOL. We were concerned that these enrichment activities would lessen the contrast between comparison and treatment groups. Consequently, we ran a sensitivity analysis removing this school from the study.

Context. There were no important contextual changes in the schools or communities in which the study was implemented. All five schools remained in the study for the entire period. In two of the three

comparison schools, and in one of the two treatment schools, there was a change of principal. The transitions were smooth in terms of acceptance of the study.

B. Impact study findings

1. Benchmark sample findings

All impact analyses were conducted on the benchmark sample (n = 491).

Primary research question (treatment effects on having ever had sexual intercourse at immediate posttest, Table IV.1). Equal percentages of youth in the treatment and comparison groups reported having sex. This was a rare event in the sample (only 2% reported having ever had sex across both the treatment and comparison conditions).

Secondary research question (treatment effects on having ever had sexual intercourse at six month followup (Table IV.1). We observed no differences between the treatment and comparison groups on having ever had sex at the six-month followup, and found no differential treatment effects for boys versus girls.

Treatment effects on other posttest and followup measures (Table IV.1). We found no statistically significant effects on any of the additional outcomes at posttest, with the exception of self-reported substance use. Treatment group youth reported lower rates of substances use than comparison youth. At the six month followup, we found no differences between the treatment and comparison condition on any of the additional non-sexual behavior outcomes.

Sex as a moderator of treatment effects (Table IV.2). We found that gender was a factor in the treatment's effect on substance use at posttest. Specifically, boys in the treatment condition had lower rates of substance use than in the comparison condition, while girls showed no differences. This was not the case at followup, nor did we find any other outcomes on which sex of participant was a factor.

2. Sensitivity analyses findings

Generally, the sensitivity analyses supported those conducted on the benchmark sample. Appendix F summarizes and contrasts results from impact analyses conducted on each matching sample. In all sensitivity samples, the effect of the treatment at posttest on substance use was statistically

significant. This was found for the benchmark and one sensitivity sample, being nonsignificant (p=.08) in the other sensitivity sample. The finding that treatment had an effect for boys but not girls was found in all three samples.

Table IV.1. Posttest and followup estimated effects examining primary, secondary, and other research questions: Treatment compared to comparison

Outcome measure	Treatment mean or % (standard deviation)	Comparison mean or % (standard deviation)	Difference	*p*-value of difference
Posttest (cohorts 1, 2, and 3)				
Behavioral measure:				
Ever had sex	2%	2%	0%	.68
Other measures				
Used substances[a]	9%	15%	-6%	.046
Internal assets: Learning	2.25 (0.61)	2.22 (0.62)	0.03	.89
Internal assets: Values	2.26 (0.54)	2.19 (0.58)	0.07	.65
Internal assets: Social	2.20 (0.52)	2.21 (0.54)	0.01	.73
Internal assets: Identity	2.28 (0.56)	2.24 (0.58)	0.04	.14
Internal assets: Total Score	2.24 (0.48)	2.21 (0.52)	0.05	.49
Cultural factor: Ethnic Identity	2.85 (0.91)	2.72 (1.01)	0.13	.09
Mental health: Anxiety	0.85 (0.66)	0.77 (0.69)	0.08	.53
Mental health: Depression	0.68 (0.62)	0.60 (0.66)	0.08	.20
Sample Size	215	276		
Followup (cohorts 1 and 2)				
Behavioral measure:				
Behavioral measure: Ever had sex	4%	1%	+3%	.19
Other measures				
Used substances[a]	20%	22%	-2%	.78
Internal assets: Learning	2.13 (0.65)	2.17 (0.67)	0.04	.96
Internal assets: Values	2.15 (0.59)	2.07 (0.61)	0.08	.21
Internal assets: Social	2.15 (0.61)	2.16 (0.58)	0.01	.36
Internal assets: Identity	2.23 (0.59)	2.19 (0.68)	0.04	.10
Internal assets: Total Score	2.16 (0.56)	2.14 (0.57)	0.02	.23
Cultural factor: Ethnic Identity	2.70 (1.02)	2.63 (1.10)	0.07	.95
Mental health: Anxiety	0.73 (0.62)	0.69 (0.72)	0.04	.26
Mental health: Depression	0.50 (0.57)	0.55 (0.69)	0.05	.61
Sample Size	146	207		

[a]Used alcohol, marijuana, or tobacco

Table IV.2. Posttest estimated effects examining sex of participant as a moderator of treatment: Treatment compared to comparison
Note. Only outcomes for which females and males differed are shown

Outcome measure	Treatment%	Comparison%	Difference	*p*-value of difference
Posttest (Cohorts 1, 2, and 3)				
Behavioral measure: Used substances[a]				
Female X Condition, *p*=.02				
Female	12%	12%	0%	.69
Male	6%	18%	-12%	.007
Sample Size	Female117 Males 99	Females 113 Males 155		

[a]Used alcohol, marijuana, or tobacco

V. Conclusion

The analysis found no significant differences on the percentage of study participants who reported ever having sexual intercourse. As noted above, this is a very young sample and very few initiated sexual activity over the course of the followup period. A longer-term followup may have captured more sexual behavior activity, which may have made it more likely to detect program effects on this outcome. Given this, we also conducted exploratory analyses to investigate whether there were impacts on outcomes that research shows are typically associated with youth having sex, which could be indicators of future early sexual behavior and teen pregnancy. The overall picture revealed that the treatment resulted in short-term but not long-term substance use reduction among males. Otherwise, we found no differences from the comparison condition on any of these outcomes on the benchmark sample. In sensitivity sample 2, we found that treatment had a positive effect on two long-term positive youth development measures-positive values and positive identity.

Limitations and discussion. Rigorous analytic approaches in applied research in small communities may unwittingly "wash out" real outcomes because of small sample sizes. This was further compounded by the fact that Cohort 3 was not included in the followup data collection (it fell outside of the time frame of the study), resulting in an even smaller sample size at followup when a number of changes over time may have become statistically significant had the sample size been larger.

Another feature of conducting research in American Indian communities is that many of these communities have been "over studied," often by researchers with little cultural sensitivity, resulting in negative perceptions of research and researchers. Nonetheless, consistency checks performed on survey data seem to indicate that those youth who did participate in the study, did respond honestly. Although the WOL program and evaluation teams were well known and respected in the target communities, some problems surfaced including the reluctance of participants and their parents to grant consent to participate in the study. Students were well aware that they would receive an incentive just for returning the consent form, even if it did not grant consent. This resulted in a lower than expected consent rate of 72% of the eligible population. An examination of potential differences between consented and non-consented groups might have revealed systematic differences though we were not able to obtain data needed for this type of analysis with the exception of age, grade, and sex of participant which were equivalent.

We opted to use positive youth development and mental health measures with published psychometric properties, however, neither of these was normed on samples that included large numbers of American Indian youth, especially from the Southwest United States. We would like to see more culturally aligned measures developed for Native youth.

VI. References

[i] Mathews TJ, Hamilton BE. Delayed childbearing: More women are having their first child later in life. NCHS data brief, no 21. Hyattsville, MD: National Center for Health Statistics. 2009: 4.

[ii] Margaret Wright. Report: State, schools lack clear strategy for tackling teen pregnancy. The Santa Fe New Mexican. May 16, 2015: 5.

[iii] http://legacy.nreppadmin.net/ViewIntervention.aspx?id=102

[iv] https://www.crimesolutions.gov/ProgramDetails.aspx?ID=235

[v] Brendtro LK, Brokenleg M, Van Bockern S. Reclaiming Youth at Risk: Our Hope for the Future. Bloomington, IN: National Educational Service; 1990.

[vi] Lerner RM, et al. Positive youth development, participation in community youth development programs, and community contributions of fifth grade adolescents: Findings from the first wave of the 4-H Study of Positive Youth Development. J Early Adolescence. 2005; 25(1): 17-71.

[vii] Developmental Assets Profile: Group Assessment User Manual. Minneapolis, MN: Search Institute; 2005.

[viii] Phinney J. The multigroup ethnic identity measure: A new scale for use with diverse groups. J Adolescent Research. 1992; 7: 172-173.

[ix] Derogatis LR, Rickels K, Rock A. The SCL–90 and the MMPI: A step in the validation of a new self–report scale. British Journal of Psychiatry, 1976; 128: 280–289.

Appendix A: Data collection efforts

Table A. Impact data collection efforts and timing

Data collection effort	Cohort 1	Cohort 2	Cohort 3
Start date of programming	09/11/12	09/25/13	09/02/14
Baseline survey	09/10/12–10/31/12	09/04/13-10/23/13	08/27/14-10/09/14
Immediate post-program follow up (posttest)	05/01/13-05/09/13	04/04/14-05/20/14	04/09/15-05/07/15
Six-month post program followup (followup)	12/04/13-01/16/14	01/21/15-02/12/15	Not Applicable

Appendix B: Implementation evaluation data collection

Implementation element	Types of data used to assess whether the element of the treatment was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Adherence: How often were sessions offered? How many were offered?	Attendance, duration, and description of each session for each participant.	Facilitators turn in attendance and fidelity logs after each session. Data are reviewed weekly and reminders are sent if data are missing.	Facilitators
Adherence: What and how much was received?	Daily attendance records	Student attendance at every session is captured in program Excel workbook	Facilitators
Adherence: What content was delivered to youth?	OAH Facilitator Logs	Facilitator logs are completed for every session	Facilitators
Adherence: Who delivered material to youth?	List of staff members hired and trained to implement program Updated training records in personnel file Background qualifications of staff members from staff applications and annual performance reviews Staff names provided on Facilitator Fidelity Logs for each session	Annual performance reviews Training records updated as training occurs	Program Coordinator
Quality: Quality of staff-participant interactions	Observations of interaction quality using protocol developed by OAH and local Evaluation Team	Random sample of 10% of classroom sessions, are selected for observation. In addition, selected out of school time sessions are selected to ensure that all types of sessions, locations, and facilitators are observed.	Independent observer member of evaluation team
Quality: Quality of youth engagement with program	Observations of interaction quality using protocol developed by OAH and local Evaluation Team	Random sample of 10% of classroom sessions, are selected for observation. In addition, selected out of school time sessions are selected to ensure that all types of sessions, locations, and facilitators are observed.	Independent observer member of evaluation team

Implementation element	Types of data used to assess whether the element of the treatment was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Quality: Stakeholder perceptions of program	Responses to interview questions developed by program Evaluation Team	Data collected annually from youth participants, teachers, principals, and parents	Independent observer member of evaluation team
Counterfactual: Experiences of comparison condition	Interviews with school personnel and selected youth	In conjunction with each of three data collection points	Evaluator
Context: Other TPP programming available or offered to study participants (both treatment and comparison)	District/school website listing all programming Interview with school counselors and teachers	Once per year Ad hoc	Facilitators and Evaluator
Context: External events affecting implementation	Interviews with school personnel, community members Review of news items	Once per year Ad hoc	Program Coordinator and Evaluator
Context: Substantial unplanned adaptation(s)	Adaptation request from school or staff, work plan, 6 month progress report, annual progress report	Annually Ad hoc	Facilitators and Evaluator

TPP = Teen Pregnancy Prevention; OAH = Office of Adolescent Health, U.S, Department of Health and Human Services.

Appendix C: Study sample

Number of:	Time period	Total sample size	Treatment sample size	Comparison sample size	Total response rate	Treatment response rate	Comparison response rate
Clusters: At beginning of study		5	2	3			
Clusters: Contributed at least one youth at baseline	Baseline	5	2	3	100%	100%	100%
Clusters: Contributed at least one youth at followup	Immediately post-programming (posttest)	5	2	3	100%	100%	100%
Clusters: Contributed at least one youth at followup	6 months post-programming (followup)	5	2	3	100%	100%	100%
Youth: In non-attriting clusters/sites at time of assignment		1,089	399	690			
Youth: Who consented		783	250	533	72%	63%	77%
Youth: Contributed a baseline survey		738	243	495	68%	61%	72%
Youth: Contributed a baseline and a posttest survey	Immediately post-programming (posttest)	619[a]	221	398	57%	55%	58%
Youth: Contributed a six-month followup survey*	6 months post-programming (followup)	359[b]	134	225	48%[c]	48% [c]	48% [c]

***** Note**. Cohorts 1 and 2 contributed to the 6-months followup time period; Cohort 3 did not. This is reflected in the lower percentages for the followup**

[a]Benchmark analysis sample after matching, n = 491 (215 treatment and 276 comparison)

[b]Benchmark analysis sample after matching, n = 353 (146 treatment and 207 comparison)

[c]Response rates calculated as the percentage of those retained among eligible individuals in Cohorts 1 and 2 only (n = 276 treatment, 470 comparison, 746 total)

Appendix D: Implementation evaluation methods

Implementation element	Methods used to address each implementation element
Adherence: How often were sessions offered? How many were offered?	The total number of sessions is a sum of the sessions captured in the Excel workbook. Average session duration is calculated as the average of the observed session lengths, measured in minutes. Average weekly frequency is calculated as the total number of sessions divided by the total number of weeks when programming was offered. *Note*. All of these calculations are also computed for type of treatment (classroom, after-school, weekend, multiday outings).
Adherence: What and how much was received?	Average number of sessions attended will be calculated as the average of the number of sessions that each student attended. Percentage of sessions attended will be calculated as the total number of sessions attended divided by the total number of sessions offered. *Note*. All of these calculations are also computed for type of treatment (classroom, after-school, weekend, multiday outings).
Adherence: What content was delivered to youth?	Total number of topics covered is the combination of the topics indicated by Facilitators in Fidelity Logs completed for each session. Proportion of topics covered will be calculated as the total number of topics completed (as indicated in Fidelity Logs) divided by total number of topics listed in the curriculum.
Adherence: Who delivered material to youth?	Total number of staff delivering the program is a simple count of staff members implementing the program as indicated in the Facilitator Logs for each session. Number and type of staff at each program site is a constant. Two facilitators at each site. Training for each staff is a constant – all staff receive the same training related to positive youth development, experiential learning cycle, facilitation and group management, as well as in required record keeping.
Quality: Quality of staff-participant interactions	An indicator of staff-participant interactions will be calculated as the % of observed interactions where the independent evaluator scored the interaction as “high quality”.
Quality: Quality of youth engagement with program	A benchmark of the quality of youth engagement is calculated as the % of sessions where the independent evaluator scored youth engagement as “moderately engaged” or higher.
Counterfactual: Experiences of counterfactual condition	The data from school personnel and students on experiences of the counterfactual at both followup points is described and a determination will be made whether these experiences have potential to impact outcomes.
Context: Other TPP programming available or offered to study participants (both treatment and counterfactual)	The data from school personnel on other TPP programming of the counterfactual throughout the study is listed in the final report.

Implementation element	Methods used to address each implementation element
Context: External events affecting implementation	All identified external events are listed in the final report.
Context: Substantial unplanned adaptation(s)	Substantial unplanned adaptations are indicated in the final report.

TPP = Teen Pregnancy Prevention.

Appendix E: Matching procedures and results

Because of the nonequivalence between conditions, we performed Exact Matching (EM) and Propensity Score Matching (PSM) procedures to create more equivalent analyses samples. We performed the matching separately for each cohort, and used the R MatchIt utility. Following the match, we merged the cohorts, and examined equivalence in two ways: among the merged data for all three cohorts and among the merged data for Cohorts 1 and 2 (the subsample used for testing the effect of condition at the six-month followup). All matching models allowed greater than one comparison participant to be matched to each treatment participant. Table E.1 summarizes the three best matching models that resulted in samples used in analyses. Our matching models using PSM model retained fewer participants (82%), but did achieve equivalence on all demographic characteristics and baseline-measured outcomes. The exception was that sensitivity sample 1 did not have equivalence on anxiety and depression. While the use of Exact Matching retained the bulk of the sample, it resulted in non-equivalence on one demographic characteristic and three outcomes. Table E.2 shows treatment versus comparison conditions in the three samples.

Table E.1. Summary of EM and PSM matching procedures

Model	Type	Demo-graphics included in match	Baseline outcomes included in match	n	Retention -Enrolled sample (n = 783)	Retention -Matching sample[a] (n = 598)	Remaining non-equivlence[b] Demo-graphics	Remaining non-equivlence[b] Baseline outcomes
Benchmark	PSM	Yes	Yes	491	63%	82%	None	None
Sensitivity 1 (four-school subsample)	PSM	Yes	Yes	438	56%	73%	None	Anxiety Depression
Sensitivity 2	Exact	Yes	No	524	67%	88%	Pueblo	Values Anxiety Depression

[a]Matching sample were those entered into the matching procedure; inclusion criteria was having pretest and posttest data (n = 619, 79% of enrolled sample) and complete data on demographic variables (n = 598, 76% of enrolled sample)

[b]Unless listed here, treatment and comparison conditions did not differ on a demographics or baseline-measured outcome variables ($p < .05$)

Table E.2. Equivalence between treatment and comparison groups

Treatment compared with comparison	Benchmark difference	*Benchmark p-value*	Sensitivity 1 difference	Sensitivity 1 *p-value*	Sensitivity 2 difference	Sensitivity 2 *p-value*
Cohorts 1, 2 3						
For post test						
Ever had sex	<1%	.73	<1%	.61	-2%	.09
Demographics	.	.	.	.	.	.
Gender (female)	-4%	.49	-5%	.27	0%	.97
Race/ethnicity: American Indian	-1%	.78	-2%	.84	-2%	.84
Race/ethnicity: Hispanic	-2%	.70	-1%	.96	1%	.77
School type: Pueblo (Navajo)	5%	.24	5%	.82	-15%	<.01
Age	-.01	.88	-.02	.65	-.02	.79
Other outcomes						
Used substances[a]	-2%	.55	-1%	.70	-2%	.25
Internal assets: Learning	.06	.27	.04	.58	.09	.06
Internal assets: Values	.08	.10	.06	.16	.11	<.01
Internal assets: Social	-.06	.10	-.07	.15	-.01	.77
Internal assets: Identity	-.07	.16	-.09	.09	-.03	.59
Internal assets: Total Score	.01	.78	-.01	.86	.05	.25
Cultural factor: Ethnic Identity	.04	.67	.03	.80	.04	.81
Mental health: Anxiety	.11	.13	.15	.03	.20	<.001
Mental health: Depression	.06	.30	.12	.02	.13	<.001
Sample size	491		438		xx	
Cohorts 1 and 2						
For follow-up						
Ever had sex	<1%	.81	<1%	.63	-1%	.32
Demographics						
Gender (female)	-6%	.26	-9%	.08	8%	.16
Race/ethnicity: American Indian	-3%	.59	-1%	.75	-3%	.65
Race/ethnicity: Hispanic	-2%	.63	-1%	.75	2%	.71
School type: Pueblo (Navajo)	-8%	.12	7%	.17	-12%	.03
Age	.02	.81	0	.93	-.02	.75
Other outcomes						

Treatment compared with comparison	Benchmark difference	*Benchmark p-value*	Sensitivity 1 difference	Sensitivity 1 *p-value*	Sensitivity 2 difference	Sensitivity 2 *p-value*
Used substances[a]	-2%	.56	-2%	.70	-5%	.33
Internal assets: Learning	.01	.73	.02	.69	.06	.24
Internal assets: Values	.07	.21	.07	.23	.09	.10
Internal assets: Social	-.07	.30	-.06	.33	-.02	.59
Internal assets: Identity	-.08	.21	-.09	.14	-.04	.49
Internal assets: Total Score	0	.78	.00	.88	.03	.38
Cultural factor: Ethnic Identity	.03	.88	.04	.81	.07	.61
Mental health: Anxiety	.11	.13	.15	.05	.19	<.01
Mental health: Depression	.06	.40	.11	.05	.10	.12
Sample size	353		300		334	

[a]Used alcohol, marijuana, or tobacco

Appendix F: Sensitivity analyses

As described in the text and in Appendix E, we created two sensitivity samples in addition to our benchmark sample. Sensitivity sample 1 was created using Propensity Score Matching, and excluded one of the five schools that are included in the other samples. Sensitivity sample 2 was created using Exact Matching.

Table F.1. Sensitivity of impact analyses using data to address questions about the effects of treatment on primary and other outcomes

Treatment compared with comparison	Benchmark difference	*Benchmark p-value*	Sensitivity 1 difference	Sensitivity 1 *p-value*	Sensitivity 2 difference	Sensitivity 2 *p-value*
Posttest (cohorts 1, 2, and 3)						
Behavioral measure						
Ever had sex	0%	.68	+1%	*.44*	+1%	.26
Other measures						
Used substances[a]	-6%	.046	-9%	*.02*	-6%	*.08*
Internal assets: Learning	0.03	.89	.02	*.92*	0.03	*.63*
Internal assets: Values	0.07	.65	.05	*.87*	0.09	*.60*
Internal assets: Social	0.01	.73	.00	*.92*	0.02	*.54*
Internal assets: Identity	0.04	.14	.01	*.34*	0.06	*.16*
Internal assets: Total Score	0.05	.49	.01	*.68*	0.05	*.53*
Cultural factor: Ethnic Identity	0.13	.09	.11	*.23*	0.04	*.88*
Mental health: Anxiety	0.08	.53	a	*.26*	0.13	*.60*
Mental health: Depression	0.08	.20	.15	*.14*	0.13	*.28*
Sample size	491		438		524	
Followup (cohorts 1 and 2)						
Behavioral measure						
Behavioral measure: Ever had sex	+3%	.19	+3%	*.14*	+1%	.17
Other measures						
Used substances[a]	-2%	.78	-.01	*.97*	0%	*.80*
Internal assets: Learning	0.04	.96	-.06	*.96*	0.02	*.62*
Internal assets: Values	0.08	.21	.04	*.30*	0.12	*.02*
Internal assets: Social	0.01	.36	-.05	*.52*	0.02	*.24*

Treatment compared with comparison	Benchmark difference	*Benchmark p-value*	Sensitivity 1 difference	Sensitivity 1 *p-value*	Sensitivity 2 difference	Sensitivity 2 *p-value*
Internal assets: Identity	0.04	.10	-.02	*.29*	0.09	*.04*
Internal assets: Total Score	0.02	.23	-.02	*.29*	0.07	*.07*
Cultural factor: Ethnic Identity	0.07	.95	.13	*.73*	0.07	*.88*
Mental health measure: Anxiety	0.04	.26	.08	*.23*	0.11	*.34*
Mental health measure: Depression	0.05	.61	.01	*.94*	0.00	*.61*
	353		300		334	

[a]Used alcohol, marijuana, or tobacco

Table F.2 Sensitivity of impact analyses examining sex as a moderator of treatment
Note. only outcomes with statistically significant differences between sexes are shown

Treatment compared with comparison	Benchmark difference	*Benchmark p-value*	Sensitivity 1 difference	Sensitivity 1 *p-value*	Sensitivity 2 difference	Sensitivity 2 *p-value*
Posttest (cohorts 1, 2, and 3)						
Behavioral: Used substances[a]						
Females	0%	*.69*	-3%	*.83*	-1%	*.68*
Males	-12%	*.007*	-16%	*.001*	-13%	*.02*
Sample size	491		438		524	

[a]Used alcohol, marijuana, or tobacco