Evaluation of Love Notes and Reducing the Risk in Louisville, KY

Final Impact Report for

University of Louisville Research Foundation

January 2016

Prepared by

Michael R. Cunningham, Ph.D.
Department of Communication, University of Louisville

Michiel A. van Zyl, Ph.D.
Kent School of Social Work, University of Louisville

Kevin Borders, MSSW, Ph.D.
School of Social Work, Spalding University

Cunningham, M. R., van Zyl, M. A., & Borders, K. W. (2016). Evaluation of Love Notes and Reducing the Risk in Louisville, Kentucky. Final Evaluation Report to the University of Louisville Research Foundation, Louisville, KY.

EVALUATION OF REDUCING THE RISK (RTR) & LOVE NOTES (LN) IN LOUISVILLE, KY: FINDINGS FROM AN INNOVATIVE TEEN PREGNANCY PREVENTION PROGRAM

I.   Introduction

As of January 2009, Kentucky ranked 8th highest in the US in teenage births, with a teenage birth rate of 51.3 per 1,000 females ages 15-19 years of age, which was significantly higher than the national rate which was 39.1 per 1,000 females. The birth rate of Non-Hispanic Black females ages 15-19 in Kentucky was even higher at 57 per 1,000.[1] Major contributing factors to high adolescent pregnancy and birthrates are engagement in high risk sexual behaviors such as having multiple partners and lack of consistent use of condoms and other forms of birth control. In Kentucky over 24% of high school students reported having had four or more partners by graduation and over 50% of sexually active students had not used a condom during their last sexual intercourse[2]. In particular, youth in the foster care system are vulnerable to the desire to form one's own family so as to counteract early childhood trauma. Studies found that by age 19, half of the young women in foster care had been pregnant and one third had given birth[3], which is 2.5 times the rate of non-foster youth (20%). Furthermore, by age 21, 71% of former foster youth became pregnant[4]. Since Louisville is a major refugee resettlement area, this vulnerable group was also targeted. Only a handful of studies have examined the impact of immigration on teen pregnancy and most of these studies only focused on Latinas[4]. To date, no studies have focused on teen pregnancy issues among refugee youth from Africa or Asia.

A.   Introduction and study overview

One recent review found that comprehensive sex education programs are effective in reducing high risk sexual behavior[5] and another confirmed that a program tested in the current study, *Reducing the Risk* (*RtR*), increased contraceptive use.[6,7] However, this *RtR* study was conducted in the 1980s. Two follow up studies[8,9] on *RtR* showed promising results, but did not meet the criteria as a quality study in the latest systematic review.[6] So, it is not certain that youth in the 21st Century will respond the same way to the curriculum. Thus, while we chose to test the 5th Edition of *RtR*[10] because it is on the HHS list of effective programs, we made adaptations to the

timing and setting of presentation of the curriculum, among other changes[11] in order to test the effectiveness in meeting the needs to today's youth.

We also tested a new approach to teen pregnancy prevention. *Love Notes* (*LN*) embeds pregnancy and disease prevention messages in a curriculum that emphasizes the importance of forming healthy relationships and avoiding intimate partner control and violence in order for individuals to reach their life goals. Studies have found that intimate partner violence (IPV) is related to sexual risk taking, inconsistent condom use, partner non-monogamy and unplanned pregnancy.[12] A focus on this destructive dynamic is not emphasized in most teen pregnancy prevention interventions. Research on an early version of *Love Notes* (*Love U2: Relationship Smarts*) with high risk youth delivered through the public school system found an impact on awareness of healthy versus unhealthy relationship patterns and reduction of verbal aggression.[13] A subsequent study with high risk youth using *LN* across two days in a community based organization found that students enjoyed the training, significantly increased their knowledge about relationships, showed a significantly lower acceptance of violence in dating relationships, and significantly increased communication and conflict management skills.[14] However, the efficacy of *LN* as a teen pregnancy prevention intervention has not been tested until now.

The purpose of this study was two-fold. First, we set out to test the efficacy of an adapted version of *RtR*, compared to a counterfactual condition, *The Power of We*. Second, we tested the efficacy for the first time of a new teen pregnancy prevention intervention, *Love Notes*, compared to the same counterfactual condition. The study was aimed at unmarried youth, ages 14-19, living in impoverished urban neighborhoods in western and southern Louisville with an emphasis on refugee and foster youth in order to understand which interventions work in the 21st Century and for which groups. This report describes the implementation and impact of each intervention on key outcomes of condom and birth control use, number of sexual partners, number who remained virgins and number of pregnancies.

B. Primary research question(s)

The primary research questions relate to two of the seven HHS Pregnancy Prevention Evidence Review outcomes[1] for each of the interventions compared to the control condition: (1) contraception use and (2) number of sexual partners.

Research Question 1: (a) Do participants in the *Reducing the Risk* intervention use condoms and other forms of birth control more often than participants in the *Power of We* control condition at a point 3 months after the conclusion of the program? (b) Do participants in the *Love Notes* intervention use condoms and other forms of birth control more often than participants in the *Power of We* control condition at a point 3 months after the conclusion of the program?

Research Question 2: (a) Do participants in the *Reducing the Risk* intervention have fewer sexual partners from the commencement of the program to 3 months after the program, compared to *Power of We* control participants? (b) Do participants in the *Love Notes* intervention have fewer sexual partners from the commencement of the program to 3 months after the program, compared to *Power of We* control participants?

C. Secondary research question(s)

The secondary research questions relate to the same two outcomes for the 6 and 12 month follow up periods.

Secondary Research Question 1: (a) Do participants in the *Reducing the Risk* intervention use condoms and other forms of birth control more often than participants in the *Power of We* control condition at a point 6 and 12 months after the conclusion of the program? (b) Do participants in the *Love Notes* intervention use condoms and other forms of birth control more often than participants in the *Power of We* control condition at a point 6 and 12 months after the conclusion of the program?

---

[1] Sexual activity including 1) initiation, 2) frequency, 3) number of sexual partners, 4) contraception use, 5) sexually transmitted infections, 6) pregnancies or 7) births.

Secondary Research Question 2: (a) Do participants in the *Reducing the Risk* intervention have fewer sexual partners from the commencement of the program to 6 and 12 months after the program, compared to *Power of We* control participants? (b) Do participants in the *Love Notes* intervention have fewer sexual partners from the commencement of the program to 6 and 12 months after the program, compared to *Power of We* control participants?

Two additional outcomes (See footnote 1) were also tested for each of the interventions compared to the control condition: (1) recent sexual activity, and (2) pregnancy.

Secondary Research Question 3: (a) Are participants in the *Reducing the Risk* intervention less likely to become pregnant if female, or cause someone else to become pregnant if male, than participants in the *PoW* control condition at a point 3, 6 and 12 months after the intervention? (b) Are participants in the *Love Notes* intervention less likely to become pregnant if female, or cause someone else to become pregnant if male, than participants in the *Power of We* control condition at a point 3, 6 and 12 months after the intervention?

Secondary Research Question 4 : Do participants in *Reducing the Risk* and *Love Notes* vs. *Power of We* differ in sexual activity at 3, 6, and 12 months after the intervention?

II. Program and comparison programming

A. Description of program as intended

Two intervention groups each received a training intervention to reduce the chances of teen pregnancy, contraction of STIs, and abusive relationships among high risk youth in the Louisville community. The theory of change was that exposure to the curriculum content would change attitudes and behavioral intentions about sexual initiation and condom/contraceptive use, and ultimately behaviors such as number of sexual partners, use of condoms/contraception and outcomes such as pregnancy and disease transmission. The 1,026 youth who were assigned to the intervention conditions were assigned to either *LN* or *RtR*.

RtR teaches skills preventing pregnancy and the spread of disease. The 16- module, 12-hour *RtR* written curriculum contains the following modules: (1) abstinence, (2) sex and protection with an emphasis on pregnancy prevention, (3) sex and protection with an emphasis on HIV prevention, (4) abstinence, refusals, using refusal skills, (5) delaying tactics, (6) avoiding high risk situations, (7, 8) getting and using protection (two modules), (9, 10, 11) three modules of

skills integration focused on knowing and talking about protection, (12) preventing HIV and other STIs, (13) HIV risk behaviors, (14) implementing protection from STI and pregnancy (including participating in a condom demonstration), (15) sticking with abstinence and protection, and (16) a final skills integration module.

Adaptations of *RtR* included updating some information about sexually transmitted infections and birth control methods to be medically accurate and changes in five exercises to increase clarity. Three primary videos were added to enhance the curriculum concepts including videos related to abstinence, human reproduction, and birth control options. Three additional videos were added to create discussion around the concepts of HIV/STIs, pregnancy, and sexual decision- making including the Scenarios USA videos named "Reflections," "The Choices We Make," and "All Falls Down" (for details on adaptations see an article[11]). The videos added three hours to the intervention for a total of 15 hours of teen pregnancy prevention (TPP) content.

LN was developed to educate participants about healthy relationships, including issues of decision-making, communication and conflict resolution, and overall safety, including the prevention of pregnancy and sexually transmitted disease[15]. The *LN* curriculum contains the following modules: sliding vs. deciding, smart love, personality and family of origin issues in relationships, safety issues, communication warning signs, healthy communication strategies, problem-solving, commitment and relationship decision-making and sexuality in close relationships including information about the success sequence for planning purposes. *LN* presents info on intimate partner violence using the Johnson multidimensional model[16] addressing issues of risk level related to dangerous behaviors of individuals as well as couple dynamics including anger management, communication and conflict resolution skills, and various types of relationship safety (emotional, physical, sexual and commitment safety).

We included all *LN* modules and key concepts of each module were covered, but streamlined the curriculum by eliminating any redundant exercises. The PowerPoint slides for *LN* helped the facilitators know what concepts to emphasize and exercises to engage in. All facilitators were familiar with all of the background material that is included in the full curriculum to help them cover all key concepts, they just did not simply read directly from the long curriculum notes that were prepared by Marlene Pearson. Four of the videos were also shown during the

sexuality modules to reinforce concepts of anatomy, abstinence, STIs, HIV/AIDs, disease prevention and contraception/pregnancy prevention. The total curriculum time was 15 hours.

Intended Delivery and Setting

The 13-module *LN* and the 16-module *RtR* curricula were both delivered across two 10-hour sessions held on two consecutive Saturdays. The 10-hour sessions included time to collect baseline (first day) and immediate post-training (second day) evaluation data as well as time for lunch and breaks. Both curricula were facilitated in group sessions (that ranged in size from 9 to 20) by certified trainers of *LN* and *RtR* with extensive experience with the facilitation of relationship and pregnancy prevention programs. They were supported by community agency staff with experience working with youth populations within the specific targeted communities. Trainers of facilitators were accessible to address any specific issues that arose. Lists of potential problematic situations and accompanying exemplar responses were compiled ahead of program offerings and shared with all facilitators.

The interventions were hosted by 23 community based organizations including eight Neighborhood Places where governmental social service agencies (e.g., child welfare, family support, public health, mental health, public school resource centers) are co-located to serve an area containing 5,000 poor children and their families[17] seven afterschool community centers (several of which were faith-based), three community schools, three centers serving refugees and immigrants, and two organizations serving foster youth.

B. Description of counterfactual condition

The 422 control group participants participated in the *Power of We (PoW)* program. This program provided training in community organizing and community building, with no focus on changing individual lifestyle or sexual behavior. Specifically, this program focused on mobilizing community members and agencies to change environmental factors affecting poor neighborhoods such as vacant property, presence of underground economies, and poor performing schools. The training includes fundamental principles and practices of community building such as (1) how to be engaged in community agencies and how to engage institutions with their communities, (2) how to identify community assets and work collaboratively, and (3) how to be part of social networks and to develop more connected local communities. It is a locally developed program that was taught by the developers from the Network Center for

Community Change (NC3). NC3 staff modified the curriculum to match the length of the *RtR* and *LN* interventions. Similar to the interventions, *PoW* includes lectures, role plays and groups discussions.

The participants attended *PoW* at the same times, for the same amount of time (15 hours of content) and at the same venues as participants in the intervention conditions. Youth received the training over two consecutive Saturdays across two, ten hour event sessions. The ten hour sessions included time to collect baseline (first day) and immediate post-training (second day) evaluation data, similar to the intervention conditions.

III. Study design

A. Sample recruitment

High-risk youth who were involved in out-of-school activities at various youth serving organizations in the poorest and most vulnerable neighborhoods in Louisville were recruited to participate in a program. To create local identification and excitement, the program was given the acronym of CHAMPS! (**C**reating **H**ealthy **A**dolescents through **M**eaningful **P**revention **S**ervices). Recruitment strategies included presentations at participating youth serving organizations, flyers posted in key gathering areas and distributed to youth program leaders, ads at youth serving organization newsletters, and presentations at youth serving organizational staff meetings to encourage staff to refer their youth clients. Interested youth contacted the project staff through one of various means provided (face to face, phone, email, website) to pre-register for a class. At that time, staff used an enrollment script to screen for inclusion/exclusion criteria for the study. If the student met the inclusion criteria, he/she was engaged in the consent process. Participants were recruited starting in the summer of 2011 and continued through March, 2014. Once the youth arrived at the agency for the program, they were organized into clusters by the researchers and these clusters were randomly assigned to each of three conditions (*LN* vs. *RtR* vs. *PoW*) for a total sample size of 1,448 in a cluster RCT. CHAMPS camps occurred 39 times with 39 clusters receiving *RtR*, 39 clusters receiving *LN* and 31 clusters receiving *PoW*.

B. Study design

Eligibility criteria for target population:

Youths were eligible to participate in the program and research sessions if they: (a) provided informed consent from their parents or guardians, (b) provided personal assent (c) were 14 to 19 years old, and (d) were affiliated with youth serving organizations, or part of a current foster youth or former foster youth alumni group.

Exclusion Criteria:

Youths were not eligible to participate in the study if they were: (a) 13 years of age or younger and 20 years or older(b) married, (c) not able to verbally participate in English, as the program and study were conducted in English, (d) had cognitive impairment that precluded them from giving assent or informed consent, (e) not able to get parental or guardian consent to participate in the study, or (f) already pregnant or a parent (since the intervention programs' aims were to prevent first pregnancies).

Random assignment process:

This study used a cluster RCT design. At the beginning of each CHAMPS! Camp, the research manager (RM) assigned each eligible participant to a cluster (depending on total number of youth at the CHAMPS! Camp, there were two groups in some of the camps and three in others). The likelihood of assignment to a treatment condition was .33 if participants reported in sufficient n numbers for three clusters (e.g., at least 30 participants), and .50 if there were only enough participants for two groups at a given camp. Seventy-nine percent of the time, there were three groups.

When creating these clusters, the RM prioritized creating a gender balance in each cluster and ensuring that all members of the household were in the same cluster. Clusters were then randomly assigned to one of the conditions (*Reducing the Risk, Love Notes*, or *Power of We*). Randomization was performed by the RM for the grant, using statistical software. Three stratification lists were created (one with only males, one with only females, and one with households that contained more than one youth); each entry in a given list was assigned to a number. This procedure ensured that there were equal numbers of households, females, and males in each group. These numbers were used to randomly assign individuals/households to clusters. The procedure by which individuals were assigned to groups (i.e., through order on a

list, or through a swapping of students across groups to achieve balance), was not consequential, both because most individuals were randomly assigned to clusters, and because the clusters were randomly assigned to condition.

The randomization was double blind, as the evaluators were blind to each condition. The result of the cluster random assignment were "released" to the facilitators on the morning of the CHAMPS! Camp and youth were told they were in a specific color group but were not told the particular condition to which they were assigned until they completed the baseline survey on the first day. Classes were run simultaneously at each site and given colors, rather than names to minimize the potential for students to determine their condition.

Consent Process:

A full consent/assent process was used. Prior to any data collection, all parents/guardians completed informed consent forms and all participants under the age of 18 completed assents which were reviewed by the IRB of the University of Louisville conveying the nature of the study, and the benefits and risks involved. There could have been some risk of discomfort in answering personal questions, and there could have been unforeseeable risks for participation. Plans were in place such that, if a participant became emotionally distressed during this study, the protocol required that he or she be referred to appropriate services. This never happened. Participants were advised that the knowledge gained may benefit future participants in this or similar training groups, as well as individuals in general through lessons learned about strengthening relationships.

C. Data collection

1. Impact evaluation

A cluster randomized controlled trial was used with six longitudinal assessment points: pre-intervention, immediate post-intervention, and then three, six, 12 and 24 months post-intervention. Baseline data were collected at the start of the first session immediately after random assignment. Follow-up data were collected for all three conditions both at the immediate conclusion of day 2 of the camp and later. Attempts were made to reach all youth participants in all three conditions at three, six, 12 and 24 months after the end of the intervention. A number of different tracking strategies to find youth for longitudinal data collection were utilized. All three study conditions completed the same measures at all points in time and were compensated

at the same levels for all parts of the study. A total of 1,448 pre-test surveys were administered at the beginning of each CHAMPS! Camp and 1,378 post-test surveys were administered at the end of each CHAMPS! Camp (immediate post-intervention), respectively, between September 2011 and March 2014. These questionnaires included demographic questions, questions about sexual behavior and outcomes and related questions that could explain sexual behavior within and across interventions. The questionnaires were given to youth in hard copy format. Each question was read out loud by a data collector to make sure that each youth was paying attention, understood all words and moved along to complete the entire questionnaire. Youth filled in the open circle that corresponded to the answer they wanted to choose using a pen or pencil.

Three months, six months, 12 months and 24 months after the end of CHAMPS! Camp youth participated in follow up data collection events. The primary way they did that was to sign up to come to a Data Daze meeting at the same agency where the initial CHAMPS! Camp was held. They were invited several weeks before the Data Daze event and were offered at least four time options each month. At the Data Daze event, youth were provided food, incentives and the questionnaire was read to the group just as it was at the pre-intervention and immediate-post intervention periods. The majority of youth completed the survey at a group Data Daze meeting. When youth were contacted about Data Daze, if it was determined that they could not attend a Data Daze, youth are given the option to come to the CHAMPS! Camp site to complete the survey, or another location near their home or school to complete the survey. This option was also given to youth who missed a scheduled Data Daze appointment. At these individual appointments, youth either had a staff person read the survey to them or completed the paper and pencil survey on their own.

The youth who had gone to college, moved out of town or could not attend a Data Daze or individualized meeting were sent the questionnaire through Survey Monkey and asked to complete the survey on-line. About 5% of youth completed the survey on-line.

At the end of the second day of CHAMPS! Camp youth received a $75 gift card for participating in the research. At the end of the three and six month follow-up survey completion youth received a $25 gift card for participating in the follow-up research. At the end of the 12- and 24-month follow-up survey completion, youth receive a $50 gift card for participating in the follow-up research. Youth also received t-shirts, backpacks, sun glasses, pens and other trinkets

with the CHAMPS! logo on them to make them feel a part of the project and to incentivize them to continue participating in the study. All youth received the same incentives regardless of when, where or how they completed the survey. No aspect of data collection differed by condition (see the table in Appendix A).

Pre-tests were administered after registration and camp orientation. During this time, the RM managed randomization for the youth. Once the surveys were administered, any youth not registered to participate was placed on the waiting list for an upcoming camp. Once the youth completed their pre-tests and were randomized to the site, the youth received their CHAMPS! drawstring bags for their respective classrooms. Data collection for the post-tests occurred immediately after the curricula on the second day of camp. Youth received their incentives immediately following the data collection. Data collection for each follow-up took place in the month of their three, six, 12, or 24 month time point. Youth were provided 60 days to complete their survey. This includes one month for them to come to a Data Daze location and one month for CHAMPS! staff to track the youth to complete the survey in an off-site location (if they did not attend a Data Daze session).

2.   Implementation evaluation

The fidelity evaluation included measures of adherence (each *LN* and *RtR* training was observed by a trained data collector), dosage (logistics staff had students sign in and out each day of each training to determine percentage who received the full dosage) and quality (from the perspective of a trained observer, from facilitator self-report and from co-facilitator observations), as well as participant engagement (from the perspective of a trained observer, each facilitator and each participant who completed questions about the session at the end of each CHAMPS! Camp) using four different measurement tools described below (See Appendix B). Program differentiation was embedded in the fact that the content of the three trainings used in the study were all different and were all observed for adherence. In addition, differential scores on *LN* and *RtR* knowledge tests, administered at the end of each CHAMPS! Camp helped to differentiate between programs.

As youth arrived at CHAMPS! Camp on days one and two, all signed in at the beginning of the day and all signed out at the end of the day. Immediately after each sign in and sign out period, staff checked off participants in an attendance database. These data were used to assess

dosage. In addition, for all of the *RtR* and *LN* sessions, a trained observer, who was a member of the data collection team, utilized the *RtR or the LN Fidelity Observation Measure* in order to further assess dosage as well as adherence to the curricula and youth engagement. In addition, the observer rated the quality of curriculum delivery using measures at the end of each of the two days *RtR* or *LN* was executed, per cohort of participants. The *Observer Quality Rating Tool* was delivered to observers as an on-line survey. This computer program was utilized in order to facilitate immediate report generation. These reports were sent to each observer and facilitator a few days after each *RtR* or *LN* delivery day. All observers were part of the evaluation and data collection team. They were trained both in the content of the curriculum and in the use of the observation tools. Observers were assigned to either *RtR* or *Love Notes.* Each one reached acceptable levels of inter-rater reliability compared to the lead fidelity team member. The facilitators also completed a similar tool to assess their own and their facilitator partner's curriculum delivery quality. The results from this *Facilitator Quality Rating Tool* were fed back to the facilitators along with the observer ratings as part of the CQI process. Someone from the implementation team discussed the results of these three measures (ratings by the observer and each facilitator) after each delivery day of *RtR* or *LN*. Strategies for improvement were built into the refreshers that occurred right before each subsequent training execution. Finally, at the end of each CHAMPS! Camp (Day 2 of the intervention), youth participants completed a survey. The *Participant Immediate Post Training Survey* contained a number of measures focused on their training experience, sexual behavior, attitudes and background information that could impact the ability of the intervention to impact outcomes. Included in the immediate post training survey were three pertinent fidelity scales that measured facilitator competence, alliance and group cohesion.

D.   Outcomes for impact analyses

The majority of the outcome measures were taken directly from the TPP Performance Measures Survey (October 6, 2011) and included the following, 1) "Have you ever had sexual intercourse?" 2) "Have you had sexual intercourse in the last 3 months?" 3) "In the past 3 months, have you had sexual intercourse without a condom?" 4) "In the past 3 months, have you had sexual intercourse without you or your partner using any of these methods of birth control?" 5) "How many different partners have you had sex with in the last 3 months?" and 6) "To the

best of your knowledge, have you ever been pregnant or gotten someone pregnant, even if no child was born?"

Table III.1. SUMMARY STATISTICS KEY BASELINE MEASURES FOR YOUTH COMPLETING CHAMPS!

Baseline Measure	Reducing the Risk (RtR)	Love Notes (LN)	Power of We (PoW)	Training F	Training P	RrR vs. PoW p	LN vs. PoW p
Demographics							
· Age (mean)	15.77	15.69	15.71	0.533	0.587	0.472	0.811
Sample size	*431*	*412*	*365*				
· Gender (% female)	63.60	64.32	62.27	0.185	0.831	0.694	0.547
Sample size	*445*	*426*	*379*				
· Race/Ethnicity(% White)	7.62	7.96	6.04	0.623	0.537	0.381	0.293
· Race/Ethnicity(% Black)	88.79	86.18	91.86	3.286	0.038	0.161	0.010
· Race/Ethnicity(%Hispanic)	3.00	4.80	2.72	1.522	0.219	0.832	0.115
· Race/Ethnicity(%Asian)	0.45	0.23	0.26	0.185	0.832	0.637	0.943
Sample size	*446*	*427*	*381*				
Primary Questions							
· Sex Without Condom Past 3 Months (%)	13.35	14.59	16.09	0.616	0.540	0.267	0.546
Sample size	*442*	*425*	*379*				
· Sex Without Birth Control Past 3 Months (%)	9.75	13.01	12.57	0.86	1.292	0.213	0.846
Sample size	*441*	*415*	*374*				
· Number of Partners Past 3 Months	2.34	2.66	2.11	0.627	0.535	0.439	0.273
Sample size	*163*	*159*	*142*				
Secondary Questions							
· Ever Pregnant (%)	0.00	0.00	0.00	-	-	-	-
Sample size	*446*	*427*	*381*				
· Pregnant in the last 3 months (%)	0.00	0.00	0.00	-	-	-	-
Sample size	*446*	*427*	*381*				
· Ever Had Sex (%)	38.69	38.63	39.47	0.037	0.964	0.818	0.806
Sample size	*442*	*422*	*380*				
· Sex in Last 3 months (%)	24.04	24.94	26.98	0.483	0.617	0.334	0.507
Sample size	*441*	*421*	*378*				

Notes: Because of ethnic group asymmetry across conditions, analyses of reported behaviors use proportion of Black *vs*. White, Hispanic & Asian Ethnic group participants as a covariate. For number of partners, many participants reported "x", "-" or left the space blank; they were coded as missing. A p, .05 is considered significant and $p \langle .15$ is considered marginal

E. Study sample

As noted above, once youth were randomly assigned to condition (*LN*, *RtR* or *PoW*) at each of the 39 CHAMPS! Camps, they completed questionnaires at immediate-post, three, six, 12 and 24 months post intervention (or control). The total number of youth with consent was 1,448 but only 1,378 remained through the entire intervention period and completed the immediate-post questionnaire, for a response rate of 95%. At the three-month follow up, 1,090 completed the questionnaire, for a response rate of 75%. At the six-month follow up, 991 completed the questionnaire for a 68% response rate, and at the 12-month follow up, 1,034 completed the questionnaire for a 71% response rate. As the tables signify, sometimes a respondent did not answer a particular question in the long questionnaire. Thus, the sample size for each analysis shifted slightly. Table III.1 shows the demographics of the participants. Table C.1a in Appendix C shows the sample flow. The Baseline sample reported in III.1 includes participants who completed CHAMPS! and who provided data in at least one follow-up session.

F. Baseline equivalence

We conducted analyses of variance and Hierarchical Linear Modeling analyses to demonstrate baseline equivalence of treatment groups. The latter analyses controlled for nesting in clusters. Participants assigned to *RtR*, *LN* and *PoW* were tested for their baseline equivalence on all of the primary and secondary outcome variables, including: (a) their frequency of use of condom and other birth control, (b) a continuous measure of their number of sexual partners (c) reports of sexual intercourse, (d) reports of pregnancy.

Secondarily, participants assigned to *RtR*, *LN* and *PoW* were tested for their equivalency on various demographic parameters including (i) age of the participant; (ii) gender of the participant, (iii) race of the participant, (iv) urban vs. suburban vs. refugee/ immigrant status of the participant; (v) birth parent vs. foster residency status of the participant, using the same procedures described above. These baseline equivalence tests were conducted both on the first assessment and at the 3, 6 and 12 month intervals. We found a small but significant race effect. More Black participants and fewer White and Hispanic participants were in the control condition than in *Love Notes*. Conversely, more White and Hispanic participants were in *Love Notes* than in the control condition. We used the proportion of Black participants to participants from other

ethnic groups as a Level 1 variable or ANCOVA covariate for all analyses. It did not affect outcomes. We also included age and gender as Level 1 variables in all HLM analyses.

Demographic data are also presented in the three, six and 12 month tables, indicating slight differences in the proportion of African and African-American participants in the control condition compared to the treatment conditions. These were appropriately addressed by using that proportion as a covariate.

G. Methods

1. Impact evaluation

Analytic sample:

The analytic samples were composed of all assigned youth who completed the survey from which the outcome data are taken. These samples involved data pooled across multiple cohorts of implementation, or across multiple study sites. Thus, the analytic samples consisted of all participants who were assigned to participate in *RtR*, *LN* or *PoW*, with a focus on the 3-month and 6-month post-intervention period. It was expected that different training sites would have different demographics, but most participants from each training were randomly assigned to one of three arbitrary clusters. As a consequence, after testing for equivalence across conditions, and planning to make appropriate corrections via inclusion of demographic variables in the analysis, it was reasonable to pool participants who received the training at different sites or within different years and cohorts, and to analyze the data using a Cluster Randomized Control Trial approach (Bloom, 2005), further details of which are presented in the Appendix.

Model specification:

The primary statistical analyses were conducted using the Bryk & Raudenbush Hierarchical Linear Modeling 7 package from SSI. Hierarchical Linear Modeling (HLM) software adjusts for the clustered nature of the data. These adjustments were needed due to youths being assigned to clusters prior to group assignment, followed by randomization of clusters to treatment conditions. In the HLM model, Level 1 specified the outcome of interest, and included Level 1 variables based on the participant's pretreatment score on the outcome variable, the probability of being assigned to a given cluster (.333 or .50 depending on the Cohort), the Cohort in which the individual participated, and the demographic covariates noted elsewhere. The Level 2 variables

were the treatments, coded as a dummy variable for Reducing the Risk (1) vs. the *Pow*er of We (0) control and a second dummy variable for the Love Notes (1) vs. the Power of We control (0). Reports of t-statistics and probabilities for impact are based on the final HLM outcome statistics using robust standard errors for the treatment variables. Reports of adjusted means for the outcome variables are based on the fitted values produced by the HLM residual file, analyzed using the Statistical Package for the Social Science (SPSS v. 21).

As the Sensitivity Analyses section explains, results for the HLM and sensitivity analyses (both with and without covariate) were extremely similar. Consequently, the non-significant sensitivity results were deleted to avoid redundancy and excessive report length.

Covariates:

The ratio of African-American and African participants to the proportion of White, Hispanic, Asian, Native American and Pacific Island participants was used as a covariate, given the results of the baseline equivalence analysis. In addition, the participant's age and gender, and his or her standing on the pre-treatment measure of the outcome variable, were all used as covariates.

Missing data approach:

Distributions of data were examined and determinations were made whether missing data for each primary variable were random or systematic. For those missing data determined to be random, the data were not replaced, because the samples possessed adequate statistical power. Systematic missing data occurred most often with questions that allowed participants to specify their own answer, rather than choose from among a fixed set of responses. For example, some participants who were not sexually active reported "0" for the number of sexual partners and some simply left the space blank. While it might have been reasonable to impute a "0" for the participants reporting no sexual contact, we chose not to do that in this o maintain maximum data integrity for the primary analysis, but did so for the sensitivity analyses reported in Appendix G.

Adjustments for multiple comparisons:

Differences were predicted *a priori* between the two interventions, *LN* and *RtR*, versus the control condition, *PoW*, at each period of assessment following program delivery for the primary and secondary outcome measures. Assessment of the impact of the two treatments and the control condition were made in the context of a Level 2 HLM regression equation, which adjusts

for overlap among the effects, implicitly controlling for comparison of the two treatments against the control group. Benjamini-Hochberg corrections are provided for the primary research questions.

Sensitivity analyses:

We conducted sensitivity analyses estimating impacts with and without covariates and cluster effects. See all Tables in Appendix E.

Analytic approach for secondary research questions:

We followed the same HLM analytic and control strategy for the secondary research questions as for the primary questions. The use of covariates, adjustments for multiple comparisons and sensitivity analysis were identical to the procedures described above.

2. Implementation evaluation

Data were analyzed using percentage calculations. There were no limitations in the implementation evaluation which was extremely thorough. See Appendix B for the data used to address implementation research questions and Appendix D for a description of the methods used to address each implementation element. Note that a list of facilitators was developed to show certification in training on pregnancy prevention and relationship education as well as experience working with youth. All facilitators had worked with youth and were given training in the curricula, attended booster sessions about how to best train the curricula and received feedback on their performance.

IV. Study findings

A. Implementation study findings

No adaptations were made during the course of the study period. No external events affected implementation (e.g. fires, disasters, bad press about TPP). All five types of fidelity were measured for the *LN* and *RtR* interventions. These five types of fidelity triangulate to demonstrate the strength of the implementation of the curricula in the current study.

Dosage was high. Ninety-three percent of youth assigned to *RtR* participated both day 1 and day 2, 94% of youth assigned to *LN* participated all of day 1 and day 2, and 98% of youth assigned to *PoW* participated all of day 1 and day 2. In *LN*, 91% of activities in the curriculum

20

were fully covered, and 4% were shortened or lengthened. In *RtR*, 93% of planned program activities were fully covered and 4.5% were partially covered.

Quality Ratings were assessed using an 11-item survey. Observers rated the quality of delivery of *LN* with a mean of 47.7 (out of 55, where 55 represents the highest quality) and observers rated the quality of delivery of *RtR* with a mean of 51.5. Facilitators rated their co-presenters using a 5-item **Partner rating** survey. Scores were high across condition, with a mean = 28.0 (out of 30) for *LN*, 27.3 for *RtR* and 27.2 for *PoW*. For **facilitator self-ratings of quality** on the 11-item survey, *LN* mean = 50 (out of 55), *RtR* rated 49 and *PoW* rated 49. On a 15-item **quality of intervention** survey (on a 75 point scale, where 75 represents the highest quality) participating youth rated *LN* mean = 66.2, *RtR* 66.1 and *PoW* 63.4.

On **observer ratings of participant engagement** for *LN*, during 92.5% of the activities most youth were rated as listening; for 7% of activities some youth seemed to be listening. Similarly, during 86% of activities, most youth were interacting, and during 13% of activities some youth were interacting. For *RtR*, during 95% of activities most youth were listening; for 4% of activities some youth seemed to be listening. Also, during 89% of activities most youth were interacting, and during 11% of activities some youth were interacting. On the **facilitator self-rating participant engagement** (2 items) for *LN*, the score was 4.4 on a 5-point scale (where 5 represents highest engagement), for *RtR* the mean score was 4.4, and for *PoW* the score was 4.35. Youth indicated on the **Facilitator Alliance Scale** a high alliance in *LN* (26.6 out of 30), *RtR* (26.6) and *PoW* (25.9). Youth indicated on the **Group Cohesion Scale** high cohesion in *LN* (33.3 out of 45), *RtR* (33.2) and *PoW* (33.4).

For **program differentiation,** we measured knowledge of the program content that was specific to each curriculum. We found that youth in *LN* scored higher in the *LN* knowledge post-test, mean = 55.4 than did *RtR* (45.4) or *PoW* (45.1), $F(2, 1352) = 58.3$, $p \langle .0001$. Similarly, youth in *RtR* scored higher on the *RtR* knowledge post-test (M = 70.0) than did youth in *LN* (64.2) or *PoW* (55.9), $F(2,1352) = 75.9$, $p \langle .0001$.

B. Impact study findings

Primary Research Question Findings: 3 months

As previously noted, the HLM analyses included a number of Level 1 controls, including each participant's standing on the baseline measures, as well as gender, age, ethnicity, cohort and probability of being assigned to treatment versus control. Because participants were excluded if they were missing any variable in each analysis, the relevant baseline metric and sample size varied. To aid interpretation, the baseline statistics and sample size, as well as the outcome statistic, is reported for each analysis in Tables IV.1-IV.3.

At the 3-month follow up, participants in *RtR* were significantly less likely than those in *PoW* to have had sex without use of birth control $(p \langle .05)$ (See Table IV.1). There also was a marginal trend for those in the *RtR* group to have had fewer sexual partners than controls $(p \langle .15)$. Additional sensitivity analyses for Number of Partners are presented in Appendix G.

There were no significant differences on the use of condoms, perhaps due to the increase in condom use in the *PoW* control compared to baseline. At the 3-month follow-up for *LN*, no effects were found on any of the primary outcomes (sex without condoms, birth control or number of sexual partners). Because there was only one effect of $(p \langle .05)$. the Benjamini-Hochberg critical value remained $(p \langle .05)$.

Secondary Research Question Findings: 3 months

Participants in *RtR* were significantly less likely at 3 months (see Table IV.1) to have ever had sex than those in *PoW (p = .03)*, and also marginally less likely to have had sex in the last three months *(p = .06)*. Additional sensitivity analyses for Ever Had Sex are presented in Appendix H, which addressed the anomalous drop in this variable in later assessments.

Participants in the *RtR* also were significantly less likely to have been pregnant or caused a pregnancy during the past 3 months $(p \langle .05)$. Participants in *LN* were marginally less likely to have been pregnant in the past three months compared to the *PoW* control *(p = .09)*. Additional sensitivity analyses for Ever Pregnant are presented in Appendix I, which addressed the anomalous drop in this variable in later assessments.

Primary Research Question Findings: 6 months

At the 6 month follow up (see Table IV.2), participants in *RtR* were significantly less likely to have had sex without birth control than those in $PoW \; (p \langle .01)$. *RtR* participants also were marginally less likely to have had sex without a condom *(p = .08)* and had marginally fewer sexual partners than those in *PoW (p = .06)*. At the 6 month follow up, participants in *LN* were significantly less likely to have had sex without a condom in the previous 3 months $(p \langle .01)$, less likely to have had sex without birth control $(p \langle .01)$ and had marginally fewer sexual partners (p = .10) compared to *PoW*. Each of the three effects that are $(p \langle .05)$ remain significant after Benjamini-Hochberg correction (p = .001, .003, .008 have adjusted critical values of .017, .033 and .05 respectively).

Table IV.1. Post-intervention HLM estimated effects of Champs! treatments at 3 months.

Measure	Reducing the Risk	Love Notes	Power of We	RtR vs. PoW	RtR vs. PoW	LN vs. PoW	LN vs. PoW
	RtR	LN	PoW	t	p	t	P
Demographics							
Age at Baseline (mean)	15.779	15.654	15.642		.117		.507
Gender (% Female)	.644	.666	.635		.823		.386
Ethnicity (%Black)	.912	.871	.929		.563		.012
Sample Size	362	342	312				
Primary Questions							
BL Sex Without Condom Past 3 Mos (%)	12.15	13.74	13.14				
3M Sex Without Condom Past 3 Mos (%)	9.95	11.11	10.57	-.370	.979	0.026	0.711
Sample Size	362	342	312				
BL Sex Without Birth Control Past 3 Mos (%)	9.14	13.17	10.65				
3M Sex Without Birth Control Past 3 Mos (%)	7.48	8.68	11.61	-1.907	.054	-1.50	.139
Sample Size	361	334	310				
BL Number of Partners Past 3 Mos (mean)	1.769	2.342	1.836				
3M Number of Partners Past 3 Mos (mean)	.678	1.144	1.009	1.430	.154	.177	.860
Sample Size	121	111	107				
Secondary Questions							
BL Ever Pregnant (%)	0.00	0.00	0.00				
3M Ever Pregnant (%)	1.39	2.03	2.55	-1.316	.188	-.618	.537

Measure	Reducing the Risk RtR	Love Notes LN	Power of We PoW	RtR vs. PoW t	RtR vs. PoW p	LN vs. PoW t	LN vs. PoW P
Sample Size	361	345	314				
BL Pregnant in Last 3 Mos (%)	0.00	0.00	0.00				
3M Pregnant in Last 3 Mos (%)	1.11	1.17	2.89	-1.960	.050	-1.683	.093
Sample Size	361	341	312				
BL Ever Had Sex (%)	38.12	36.95	38.85				
3M Ever Had Sex (%)	30.94	34.31	35.99	-2.173	.030	-.467	.634
Sample Size	362	341	314				
BL Sex in Last 3 months (%)	23.27	23..01	25.64				
3M Sex in Last 3 months (%)	18.56	20.65	23.72	-1.880	.060	-.882	.378
Sample Size	361	339	312				

Notes: HLM analysis used Baseline outcome, age, gender, ethnicity, cohort and probability of assignment to cluster as Level 1 variables, and Treatment as Level 2 variables. For number of partners, many participants reported "x", "-" or left the space blank; they were coded as missing. A $p \langle =.05$ is considered significant

Table IV.2 Post-intervention HLM estimated effects of Champs! treatments at 6 months.

Measure	Reducing the Risk RtR	Love Notes LN	Power of We PoW	RtR vs. PoW t	RtR vs. PoW p	LN vs. PoW t	LN vs. PoW P
Demographics							
· Age at Baseline (mean)	15.73	15.69	15.62		.402		.842
· Gender (% Female)	64.04	64.06	63.79		.974		.965
· Ethnicity (%Black)	91.80	88.13	92.76		.497		.037
Sample Size	317	320	290				
Primary Questions							
· BL Sex Without Condom Past 3 Mos (%)	13.56	12.81	13.79				
· 3M Sex Without Condom Past 3 Mos (%)	12.33	9.69	16.55	-1.739	.082	-2.690	.008
Sample Size	317	320	290				
· BL Sex Without Birth Control Past 3 M (%)	11.04	11.18	12.24				
· 3M Sex Without Birth Control Past 3 M (%)	9.15	8.31	17.48	-2.982	.003	-3.392	.001
Sample Size	317	313	286				
· BL Number of Partners Past 3 Mos	2.18	1.96	1.82				
· 3M Number of Partners Past 3 Mos	.84	.77	1.77	-1.892	.060	-1.652	.100
Sample Size	88	84	79				
Secondary Questions							
· BL Ever Pregnant (%)	0.00	0.00	0.00				
· 6M Ever Pregnant (%)	2.19	.93	3.45	-.939	.348	-2.049	.041
Sample Size	319	322	290				
· BL Pregnant in Last 3 Mos (%)	0.00	0.00	0.00				

Measure	Reducing the Risk RtR	Love Notes LN	Power of We PoW	RtR vs. PoW t	RtR vs. PoW p	LN vs. PoW t	LN vs. PoW P
· 6M Pregnant in Last 3 Mos (%)	1.89	1.88	2.77	-.506	.613	-.441	.659
Sample Size	317	318	289				
· BL Ever Had Sex (%)	40.38	36.99	38.49				
· 6M Ever Had Sex (%)	18.93	18.89	27.68	-1.655	.098	-2.345	.019
Sample Size	317	319	291				
· BL Sex in Last 3 months (%)	24.68	22.08	25.77				
· 6M Sex in Last 3 months (%)	23.42	19.24	27.84	-1.367	.172	-2.236	.026
Sample Size	316	317	291				

Notes: HLM analysis used Baseline measure of outcome, age, gender, ethnicity, cohort and probability of assignment to cluster as Level 1 variables, and Treatment as Level 2 variables. For number of partners, many participants reported "x", "-" or left the space blank; they were coded as missing. A $p \langle =.05$ is considered significant

Secondary Research Question Findings: 6 months

Participants in *RtR* were marginally less likely at 6 months (see Table IV.2) to report ever having had sex compared to those in *PoW* (p = .10), but did not differ on pregnancy or sex during the past three months. By contrast, participants in *LN* were significantly less likely to have ever been pregnant or caused a pregnancy *(p = .04)*, significantly less likely to have ever had sex *(p = .02)* and significantly less likely to have had sex in the last 3 months (p = .03) than participants in *PoW*.

Primary and Secondary Research Findings: 12 months

At the 12-month follow up, there were no significant differences between either *RtR* or *LN* on the outcome measures of having had sex without a condom, having sex without birth control, and having fewer sexual partners or pregnancy compared to the *PoW* group (see Table IV.3). Because no effects were $(p \langle .05)$, there were not Benjamini-Hochberg corrections.

There were no significant differences at the 12 month follow-up between *RtR* and *PoW*, or *LN* and *PoW* in likelihood of having sex, pregnancy or causing a pregnancy (See Table IV.3)

Table IV.3 Post-intervention HLM estimated effects of Champs! treatments at 12 months.

Measure	Reducing the Risk RtR	Love Notes LN	Power of We PoW	RtR vs. PoW t	RtR vs. PoW p	LN vs. PoW t	LN vs. PoW P
Demographics							
· Age at Baseline (mean)	15.66	15.67	15.61		.793		.672
· Gender (% Female)	65.40	65.96	62.08		.670		.492
· Ethnicity (%Black)	91.50	85.54	94.63		.474		.0001
Sample Size	341	322	298				
Primary Questions							
· BL Sex Without Condom Past 3 Mos (%)	12.02	14.76	13.42				
· 3M Sex Without Condom Past 3 Mos (%)	13.20	13.86	14.43	-.733	.464	-.955	.340
Sample Size	341	322	298				
· BL Sex Without Birth Control Past 3 M (%)	9.23	12.00	12.46				
· 3M Sex Without Birth Control Past 3 M (%)	13.20	13.86	14.43	-1.250	.212	-.566	.571
Sample Size	336	325	297				
· BL Number of Partners Past 3 Mos	2.17	2.30	1.79				
· 3M Number of Partners Past 3 Mos	1.12	1.02	1.09	-.311	.756	-1.151	.251
Sample Size	107	108	95				
Secondary Questions							
· BL Ever Pregnant (%)	0.00	0.00	0.00				
· 12M Ever Pregnant (%)	7.65	4.17	5.00	1.233	.218	-.606	.545
Sample Size	340	335	300				
· BL Pregnant in Last 3 Mos (%)	0.00	0.00	0.00				
· 12M Pregnant in Last 3 Mos (%)	5.29	2.99	4.35	.630	.529	-.680	.497
Sample Size	340	335	299				
· BL Ever Had Sex (%)	37.72	39.46	39.33				
· 12M Ever Had Sex (%)	38.30	43.07	39.00	-.459	.646	.527	.598
Sample Size	342	332	300				
· BL Sex in Last 3 months (%)	23.98	24.16	26.19				
· 12M Sex in Last 3 months (%)	20.41	24.42	23.21	-1.065	.287	.255	.799
Sample Size	392	385	336				

Notes: HLM analysis used Baseline outcome, age, gender, ethnicity, cohort and probability of assignment to cluster as Level 1 variables, and Treatment as Level 2 variables. For number of partners, many participants reported "x", "-" or left the space blank; they were coded as missing. A $p \langle =.05$ is considered significant

V. Conclusion

The implementation analyses showed that the five measures of fidelity were strong. Those receiving the full dosage of each intervention, the adherence to the two intervention curricula, and the quality of delivery of each intervention were all very high. The youth engagement across four measures, including two completed by youth, was very high and the differentiation between interventions was verified via observations and tests of knowledge post-intervention. Thus, any impact of an intervention on risky sexual behavior or sexual outcomes post- intervention can

reliably be attributed to that intervention. Furthermore, while the control group received the full dose of a high–quality and engaging training, *PoW* did not include information on healthy relationships or sex education and is thus a viable control condition.

Individual Program Impact

The impact evaluation showed that while there was only a trend in reducing pregnancy rates by the 3-month follow up for participants in *Love Notes*, by six months, youth in *Love Notes* were more likely to use both condoms and other forms of birth control, were less likely to have ever had sexual intercourse, were less likely to have had sex in the last three months and were less likely to get pregnant or get another person pregnant. There was also a trend for having fewer sexual partners. However, this effect did not hold at the 12 month follow up. Such favorable results in challenging participant populations suggest that *Love Notes* can be added to the inventory of evidence-based programs, although we will offer suggestions for improvement below.

The impact evaluation also offered evidence that *Reducing the Risk* had a positive impact, as well. Youth in *Reducing the Risk* were less likely to have ever had sexual intercourse at the 3-month mark and were more likely to have used birth control when they did have sex to have fewer pregnancies. There was also a trend for fewer sexual partners. By 6- months, the result for more birth control use held up and three trends for less engagement in sexual intercourse ever, fewer partners and more use of condoms was also found. But, like *Love Notes,* these results did not extend to the 12 month follow up period. Those favorable results can justify the use of *Reducing the Risk,* particularly to focus on short term change. Other suggestions will be made below.

A particularly important finding from this investigation was that both programs produced favorable outcomes.

Program-Specific Outcomes and Enhancements

In some ways, our results fit the theory of change underlying each of these different curricula. For example, *Reducing the Risk* has many modules focused on helping youth think about and practice ways to avoid sexual situations, get out of sexual situations and delay the onset of sexual engagement.

By contrast, *Love Notes* spends several modules helping youth think through their long term life goals and plans and how sex, pregnancy, parenthood and getting involved with an abusive partner could derail those goals. As a part of that emphasis is an introduction to the "success sequence" which helps youth see the benefits of completing a high school education, going on to college or receiving other sorts of job training or mastering a technical skill, so that they can be self-sufficient before finding a mate and then having children. So, while one strategy for delaying pregnancy is abstinence, it is not the only strategy. And *LN* youth did not receive the detailed practice in fending off advances the way that the *Reducing the Risk* participants did. But they were particularly cued into the importance of reducing events that could derail their success. Thus, it is no surprise use of birth control and condoms, as well as lower rates of pregnancy were highest for those youth in this curriculum.

Exposure and Implementation Duration

Our approach involved delivering both interventions in a short time period. That ensured exposure to the full dosage of a curriculum with high fidelity. That approach successfully impacted behavior and outcomes, especially at the 3-month and 6-month assessments. We must, however, recognize that the intense exposure approach executed over two weekends may not be strong enough to reduce all forms of risky behavior, especially all of the way to the 12-month follow-up.

It is possible that it may work better to implement the intervention over a longer time period, so that the information can sink in and supports for decisions about sex can be secured. Previous research on *Reducing the Risk* consistently showed a positive outcome on delaying intercourse. The current study found that effect at 3 months. Thus, we replicated that effect. However, two previous studies[8-9] showed a marked effect on that outcome up to 18 months out. In addition, while previous research found enhanced STD and pregnancy prevention behavior (e.g. use of condoms and other forms of birth control) at the 18 month mark, our study only found the effect at the 3 and 6 month marks and not at the 12 month mark.

Perhaps the addition of booster sessions would have led to a replication of previous studies. A future study should compare implementation over a longer period versus implementation over a shorter period (with and without intensive boosters every one to three months) to see which delivery method works best to impact not only delays in sexual initiation but also use of

contraceptives. *Love Notes* also has typically been administered over longer periods of time. The fact that we had positive outcomes at 6 months despite the brief intervention period, shows the potential for *Love Notes* to be a viable intervention in reducing risky sex and pregnancy in high risk youth. Further research, like that suggested for *Reducing the Risk,* should be conducted with *Love Notes* to examine the difference in delivery timing and boosters on outcomes.

Benefits of Intensive Exposure

Although we have expressed concerns about program duration, it is important to note that the intense exposure approach used in this study has several advantages. (1) The intensive approach is highly efficient in terms of personnel, logistics, travel time, and facilities requirements and costs. Our small program staff was able to deliver two sets of interventions on 39 occasions, for a total of 78 completed program days, at 23 different community sites. We simply did not have the program staff, and could not have the secured the community facilitators' cooperation, or gained extended access to their settings, to deliver so many program days for 12 to 14 weeks each outside of a school setting. The local school system did not have the time to devote 15 hours to execute an evidence based curriculum during the school day. (2) The intensive approach allowed the program staff to maintain effective oversight and ensure fidelity for each program delivery. (3) The intensive approach is effective in securing participant engagement and retention. We retained 95% of participants from the first weekend to the second weekend, ensuring that virtually all participants received the full dose of the curricula to which they were exposed. Programs with longer duration typically suffer from higher attrition rates and greater student boredom and disengagement, and (4) The intensive exposure approach captures the participants' full attention for two full days, thereby preventing compartmentalization of the program as 'just one more class' to be endured for an hour per week. Providing information in a compressed fashion can be a benefit. Major changes, challenges and risks are occurring in the lives of youth at all times, so stretching out the presentation across 12 to 14 weeks may mean that it comes too late for some program participants. The relative strengths and weaknesses of intensive exposure versus extended duration program delivery will need systematic testing, along with the use of boosters in future research as suggested above.

Booster Sessions

Adding booster sessions either via text messages or during short sessions together may be advisable in scale-up projects using these curricula. Booster sessions may help to ensure that youth remember all of the information that they were exposed to during the intervention and can keep the principles and facts in the forefront of their minds.

We also are mindful that asking our participants to complete research questionnaires at 3, 6 and 12 months may have served as boosters, because youth were reminded of some of the material (through the knowledge test) and the goal of staying abstinent, reducing risky sexual behavior and avoiding pregnancy and disease. In hindsight, a 9-month follow up survey might have been helpful in this regard. It is conceivable that a 9-month survey would have served as a booster and enhanced the outcomes at 12 months.

Lessons Learned

While the current cluster randomized control trial was implemented well and found some interesting effects concerning two programs compared to a control, there were some limitations. (1) Although shortening the intervention delivery time increased fidelity, dosage and effects up to six months after the interventions, there may have been a downside. That is, the concentrated intervention may have been powerful for the first six months post-treatment, but then lost potency after six months, as indicated by the limited effects at 12 months.

(2) We felt it was important for randomization and for the youth to have a meaningful experience in the control condition. As a consequence, we chose not to gather data on a group of youth with the same poverty levels from the same vulnerable neighborhoods who were not engaged in any activity. In hindsight, we are concerned that the control group gained self-esteem or other attributes that influenced their engagement in risky sexual behaviors. We now regret that we did not have a wait list control group who responded to the questionnaires, to see how well the interventions compared against a control involving no uplifting experience. In a future study, we would like to include such a group.

(3) We included two very disenfranchised minority populations in the study, (foster youth and refugee youth). While this was a strength of the study because we could engage these youth in meaningful interventions or experiences, the sample sizes of these groups were small. We will have pilot data for future research on the impact of an intervention aimed at refugees and foster

youth, but we regret that we lacked the statistical power to determine if these populations respond differentially to one or the other treatment.

Although we have endeavored to learn from our experience, we suffered no major setbacks and were gratified by both the process and the outcomes of the project.

Parting Recommendations

We regard this project as highly successful. Both the *Love Notes* and the *Reducing the Risk* were delivered with fidelity, and were found to significantly impact important outcomes. Awareness of the similarities and differences between the *Love Notes* and the *Reducing the Risk* highlights an additional issue. Rather than focus on an evidence-based program as an indivisible whole, future research should examine both the common and unique content elements in evidence-based curricula and programming and (a) test the impact of each component ingredient on specific outcomes (e.g., number of sexual partners, use of condoms), and (b) test hybrid programs using modules from different programs to determine the best mix of ingredients, in order to effectively impact all seven outcomes delineated by HHS as critical for adolescent health and prevention of teen births.

Despite these limitations, this study adds to the growing literature on "what works" to reduce risky teen sexual behavior. This adaptation of *Reducing the Risk* had an impact on more than one type of risky sexual behavior using a more condensed delivery method. And, exposing youth to a heavy dose of life planning, healthy relationship and violence prevention material in the context of teen pregnancy prevention was also successful in reducing risky sexual behavior. This study of *Love Notes* contributes to public health's search for comprehensive programming to increase adolescent health across multiple areas.[18]

Addressing life planning and more than one high risk behavior may be more cost effective both in terms of time and expenditures in enhancing positive youth development and reducing maladaptive behavior.

VI. References

[1] Kost K, Henshaw S. U.S. teenage pregnancies, births and abortions, 2010: *National and state trends by age, race and ethnicity*. New York, NY Guttmacher Institute, 2014.

[2] Centers for Disease Control and Prevention: Youth Risk Behavior Surveillance—United States, 2011. MMWR 2012; 61: pp. 1-162.

[3] Dworsky A, Courtney ME. The risk of teenage pregnancy among transitioning foster youth: Implications for extending state care beyond age 18. *Child & Youth Serv Rev.*2010; 32: 1351-1356.

[4] Santelli JS, Abma J, Ventura S, Lindberg L, Morrow B, et al. Can changes in sexual behaviors among high school students explain the decline in teen pregnancy rates in the1990s? *J of Adol Health.* 2004; 35: 80-90.

[5] Chin HB, Sipe TA, Elder R, et al. The effectiveness of group-based comprehensive risk-reduction and abstinence education interventions to prevent or reduce the risk of adolescent pregnancy, human immunodeficiency virus, and sexually transmitted infections: Two systematic reviews for the Guide to Community Preventive Services. *Am J Prev Med.* 2012; 42: 272-294.

[6] Goesling B, Colman S, Trenholm C, Terzian M, Moore K. Programs to reduce teen pregnancy, sexually transmitted infections, and associated sexual risk behaviors: A Systematic review. *J of Adol Health*. 2014; 54(5): 499-507.

[7] Kirby DR, Barth RN, Leland N, Fetro JV. Reducing the Risk: Impact of a new curriculum on sexual risk-taking. *Fam Plan Persp.* 1991; 23(6): 253-263.

[8] Hubbard BM, Giese ML, Rainey J. A replication of Reducing the Risk, a theory-based sexuality curriculum for adolescents. *J of School He.* 1998; 68(6): 243-247.

[9] Zimmerman RS, Cupp PK, Donohew L, Sionean C, Feist-Price S, Helme D. Effects of a school-based, theory-driven HIV and pregnancy prevention curriculum. *Persp on Sex Repro Health.* 2008; 40(1): 42-51.

[10] Barth, R. *Reducing the Risk: 5th Edition.* Scotts Valley, CA. ETR. 2011.

[11] Langley CN, Barbee AP, Antle BF et al. Enhancement of Reducing the Risk for the 21st Century: Improvement to a curriculum developed to prevent teen pregnancy and STIs. *Am J of Sex Ed.* 2015; 10(3): 40-69.

[12] Coker, AL. Does physical intimate partner violence affect sexual health? *Trauma, Viol & Abuse.* 2007; 8(2): 149-177.

[13] Adler-Baeder F, Kerpelman J, Higginbotham B, Schramm D, Paulk A. The impact of relationship education on adolescents from diverse backgrounds. *Family Relations.* 2007; 56: 291-303.

[14] Antle BF, Sullivan DJ, Dryden AA, Karam EA, Barbee AP. Promoting healthy relationships among high risk youth. *Child &Youth Ser Rev.* 2011; 33 (1): 173-179.

[15] Pearson M. *Love Notes*. Berkeley, CA. The Dibble Institute for Marriage Education, 2011.

[16] Johnson MP, Leone JM. The differential effects of intimate terrorism and situational couple violence: Findings from the national violence against women survey. *J of Fam Issues.* 2005; 26: 322-349.

[17] Barbee AP, Antle BF. Cost effectiveness of an integrated service delivery model as measured by worker retention. *Child & Youth Serv Rev.* 2011. 33: 1624-1629.

[18] Kagesten A, Parekh J, Tuncalp O, Turke S, Blum RW. Comprehensive adolescent health programs that include sexual and reproductive health services: A systematic review. *Am J of Public Health.* 2014; 104: e23-e36.

Appendix A: Data collection efforts

Table A.1. Data collection efforts used in the impact analysis of Love Notes and Reducing the Risk and timing (mo/yr)

Data collection effort	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Cohort 6	Cohort 7
Start date of programming	09/11	10/11	10/11	11/11	12/11	01/12	02/12
Baseline survey	09/11	10/11	10/11	11/11	12/11	01/12	02/12
Immediate post-Test	09/11	10/11	10/11	11/11	12/11	01/12	02/12
3-month follow-up	12/11	01/12	01/12	02/12	03/12	04/12	05/12
6-month follow-up	3/12	04/12	04/12	05/12	06/12	07/12	08/12
12-month follow-up	9/12	10/12	10/12	11/12	12/12	01/13	02/13
24-month follow-up	9/13	10/13	10/13	11/13	12/13	01/14	02/14

Data collection effort	Cohort 8	Cohort 9	Cohort 10	Cohort 11	Cohort 12	Cohort 13	Cohort 14
Start date of programming	03/12	03/12	04/12	04/12	05/12	06/12	07/12
Baseline survey	03/12	03/12	04/12	04/12	05/12	06/12	07/12
Immediate post-Test	03/12	03/12	04/12	04/12	05/12	06/12	07/12
3-month follow-up	06/12	06/12	07/12	07/12	08/12	09/12	10/12
6-month follow-up	09/12	09/12	10/12	10/12	11/12	12/12	01/13
12-month follow-up	03/13	03/13	04/13	04/13	05/13	06/13	07/13
24-month follow-up	03/14	03/14	04/14	04/14	05/14	06/14	07/14

Data collection effort	Cohort 15	Cohort 16	Cohort 17	Cohort 18	Cohort 19	Cohort 20	Cohort 21
Start date of programming	07/12	08/12	08/12	09/12	10/12	10/12	11/12
Baseline survey	07/12	08/12	08/12	09/12	10/12	10/12	11/12
Immediate post-Test	07/12	08/12	08/12	09/12	10/12	10/12	11/12
3-month follow-up	10/12	11/12	11/12	12/12	01/13	01/13	02/13
6-month follow-up	01/13	2/13	2/13	3/13	04/13	04/13	05/13
12-month follow-up	07/13	8/13	8/13	9/13	10/13	10/13	11/13
24-month follow-up	07/14	8/14	8/14	9/14	10/14	10/14	11/14

Data collection effort	Cohort 22	Cohort 23	Cohort 24	Cohort 25	Cohort 26	Cohort 27	Cohort 28
Start date of programming	12/12	02/13	03/13	03/13	04/13	04/13	05/13
Baseline survey	12/12	02/13	03/13	03/13	04/13	04/13	05/13
Immediate post-Test	12/12	02/13	03/13	03/13	04/13	04/13	05/13
3-month follow-up	03/13	05/13	06/13	06/13	07/13	07/13	08/13
6-month follow-up	06/13	08/13	09/13	09/13	10/13	10/13	11/13
12-month follow-up	12/13	02/14	03/14	03/14	04/14	04/14	05/14
24-month follow-up	12/14	02/15	03/15	03/15	04/15	04/15	05/15

Data collection effort	Cohort 29	Cohort 30	Cohort 31	Cohort 32	Cohort 33	Cohort 34	Cohort 35
Start date of programming	06/13	07/13	08/13	09/13	10/13	10/13	11/13
Baseline survey	06/13	07/13	08/13	09/13	10/13	10/13	11/13
Immediate post-Test	06/13	07/13	08/13	09/13	10/13	10/13	11/13
3-month follow-up	09/13	10/13	11/13	12/13	01/14	01/14	02/14
6-month follow-up	12/13	01/14	2/14	3/14	04/14	04/14	05/14
12-month follow-up	06/14	07/14	8/14	9/14	10/14	10/15	11/14
24-month follow-up	06/15	07/15	8/15	9/15	10/15	10/16	11/15

Data collection effort	Cohort 36	Cohort 37	Cohort 38	Cohort 39
Start date of programming	01/14	02/14	03/13	03/13
Baseline survey	01/14	02/14	03/13	03/13
Immediate post-Test	01/14	02/14	03/13	03/13
3-month follow-up	04/14	05/14	06/13	06/13
6-month follow-up	07/14	08/14	09/13	09/13
12-month follow-up	01/15	02/15	03/14	03/14
24-month follow-up	01/15	02/16	03/15	03/15

Appendix B: Implementation evaluation data collection

Table B.1. Data used to address implementation research questions

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Adherence: How often were sessions offered? How many were offered?	All sessions offered were captured in program records and performance measure reporting system (PMRS)	All sessions delivered were captured in program records and PMRS	Logistics staff
	Length (number of minutes) of program sessions captured in *LN* and *RtR* Observation Tool	All *LN* and *RtR* sessions observed using *LN* or *RtR* Observation Tool	Data Collectors
Dosage: What modules were attended and how many days of curriculum content was received?	Sign in and sign out sheets each day of CHAMPS! Camp to show attendance at each day of camp which is a form of dosage.	Day 1 and day 2 of every CHAMPS! Camp utilized sign in and sign out sheets to show how much material youth were exposed to across the 2 days of camp	Logistics staff
Adherence: What content was delivered to youth?	Number of "activities" covered captured in *LN* and *RtR* Observation Tool	All *LN* and *RtR* sessions observed using *LN* or *RtR* Observation Tool	Data Collectors
Adherence: Who delivered material to youth?	List of facilitators hired and trained to implement program Background qualifications of facilitators from applications List of 2 facilitators for each *LN* and each *RtR* CHAMPS! Camp	Data on all facilitators were available to program staff	Program staff
Quality: Quality of training delivery	Observer Assessment Tool: Questions regarding delivery	After 86% of CHAMPS! Camp days, observers completed the Observer Assessment Tool	Data Collectors completed tool
Quality: Quality of training delivery	OAH Facilitator Self-Assessment Tool: Questions regarding delivery	After 91% of CHAMPS! Camp days facilitators completed on-line Facilitator Self-Assessment Tool	Logistics staff sent survey to facilitators
Quality: Quality of training delivery	OAH Co-Facilitator Assessment Tool: Questions regarding partner delivery	After 80% CHAMPS! Camp days facilitators completed Co-Facilitator Assessment Tool	Logistics staff send survey to facilitators regarding partner delivery
Quality: Quality of training delivery	Participant Satisfaction Tool: Questions regarding facilitator delivery	At the end of second day of each CHAMPS! Camps, youth completed questions about satisfaction with facilitator quality of delivery	Data Collectors read surveys out loud to youth participants
Quality: Quality of youth engagement with program	OAH Facilitator Self-Assessment Tool: Questions regarding youth engagement	After 80% of CHAMPS! Camp days facilitators completed on-line Facilitator Self-Assessment Tool	Logistics staff send survey to facilitators
Quality: Quality of youth engagement with program	*LN* and *RtR* Observer Assessment Tool: 2 Questions regarding the youth engagement	All *LN* and *RtR* sessions observed using Observer Assessment Tool	Data Collectors regarding each facilitator's delivery

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Quality: Quality of youth engagement with program	Facilitator Alliance Scale Group Cohesion Scale	At the end of second day of each CHAMPS! Camps, youth completed both Facilitator Alliance and Group Cohesion Scales	Data Collectors read surveys to youth participants
Program Differentiation: Between *LN*, *RtR* and *PoW*	Immediate Post-Camp *RtR* Knowledge Test Immediate Post Camp *LN* Knowledge Test	Youth took the *LN* and the *RtR* test post Camp regardless of their condition (*LN vs. RtR vs. PoW*)	Youth participants
Counterfactual: Experiences of comparison condition	All sessions offered were captured in program records	All sessions delivered were captured in program records	Logistics staff
Counterfactual: Experiences of comparison condition	Length (number of minutes) of program sessions captured in observations	30% of *PoW* sessions observed	Data Collectors
Counterfactual: Experiences of comparison condition	Sign in and sign out sheets	All youth signed in and out of *PoW*	Logistics staff
Counterfactual: Experiences of comparison condition	OAH Facilitator Self-Assessment	Facilitators completed Tool for 90% of sessions	Logistics staff; Data Collectors gave survey
Counterfactual: Experiences of comparison condition	· Participant Satisfaction Tool · Facilitator Alliance Scale · Group Cohesion Scale	At the end of second day of each CHAMPS! Camps, youth completed questions about satisfaction with facilitator quality of delivery	Youth participants
Context: Other TPP programming available or offered to study participants (both intervention and comparison)	Participant questionnaire: Questions regarding exposure to other TPP programming	Pre-, immediate post-, 3-,6-,12-,and 24-month follow up periods asked about exposure to other TPP programming	Data Collectors/Evaluators
Context: External events affecting implementation	Surveyed news stories	Throughout the grant period	Program Staff
Context: Substantial unplanned adaptation(s)	*LN* and *RtR* Observation Tool	All *LN* and *RtR* sessions were observed	Data Collectors

Notes: TPP = Teen Pregnancy Prevention

Appendix C: Study sample

Table C.1a. Cluster and youth sample sizes by intervention status – cluster designs

Number of:	Time period	Total Sample Size	Intervention Sample Size *LN*	Intervention Sample Size *RtR*	Comparison Sample Size	Total response rate	Intervention response rate	Comparison response rate
Clusters: At beginning of study		109	39	39	31	N/A	N/A	N/A
Clusters: Contributed at least one youth at baseline	Baseline	109	39	39	31	100%	100%	100%
Clusters: contributed at least one youth at follow-up	Immediately post-programming	109	39	39	31	100%	100%	100%
Clusters: Contributed at least one youth at follow-up	3-months post-programming	109	39	39	31	100%	100%	100%
Clusters: Contributed at least one youth at follow-up	6-months post-programming	109	39	39	31	100%	100%	100%
Clusters: Contributed at least one youth at follow-up	12-months post-programming	109	39	39	31	100%	100%	100%
Clusters: Contributed at least one youth at follow-up	24-months post-programming	109	39	39	31	100%	100%	100%
Youth: In non-attriting clusters / sites at time of assignment		1448	511	517	422	N/A	N/A	N/A
Youth: Who consented		1466*						
Youth: Contributed a baseline survey		1448	511	515	422	100%	99.6% 100%	100%
Youth: Contributed a follow-up survey	Immediately post-programming	1378	484	481	413	95%	94% 93%	98%
Youth: Contributed a follow-up survey	3-months post-programming	1090***	367	386	337	75%	72% 75%	80%
Youth: Contributed a follow-up survey	6-months post-programming	991	345	338	308	68%	67% 65%	73%
Youth: Contributed a follow-up survey	12-months post-programming	1034	405	411	352	71%	79% 80%	83%
Youth: Contributed a follow-up survey**	24-months post-programming	638/1060	215	233	190	60%	57% 62%	62%

Notes: *18 participants gave initial consent but withdrew from the study before random assignment to condition
**Still collecting 24-months post-programming data
*** Since the analyses include (a) pre-intervention measures, (b) post-intervention measure (c) gender (d) age and (e) ethnicity meant that there are missing data in the HLM analyses. Tables in section IV include actual numbers of subjects for each analysis. This table focuses on number of subjects enrolled and who completed follow up surveys

Appendix D: Implementation evaluation methods

Table D.1. Methods used to address implementation research questions

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 the sum of the sessions captured in the program files. Average session duration is calculated as the average of the observed session lengths, measured in minutes and reported as hours and minutes. Only those youth that attended both full days of CHAMPS! Camp were counted as receiving the full dosage.
Adherence: What and how much was received?	This calculation was cross checked with the sign in, sign out sheets for each day of CHAMPS! Camp for each intervention and control group.
Adherence: What content was delivered to youth?	Total number of activities covered fully and total number of activities covered but shortened.
Adherence: Who delivered material to youth?	100% of facilitators delivering Love Notes and 100% of facilitators delivering Reducing the Risk were trained and coached in their curricula. All had experience working with youth.
Quality: Quality of delivery	Eleven questions on the observer assessment tool regarding delivery quality were added together and mean scores for each facilitator were calculated and the over-all mean for each treatment condition (*LN* and *RtR*) were calculated. Range in scores from 11 to 55.
Quality: Quality of delivery	Eleven questions on the facilitator self-assessment tool regarding delivery quality were added together and mean scores for each facilitator were calculated and the over-all mean for each treatment condition (*LN* and *RtR*) were calculated. Range in scores from 11 to 55.
Quality: Quality of delivery	Five questions on the Partner Assessment Tool regarding delivery quality were added together and mean scores for each facilitator were calculated and the over-all mean for each treatment condition (*LN* and *RtR*) were calculated. Range in scores from 5 to 25.
Quality: Quality of delivery	Fifteen questions on the Participant Satisfaction Tool regarding delivery quality were added together and mean scores for each Camp was calculated and the over-all mean for each treatment condition (*LN* and *RtR*) were calculated. Range in scores from 15 to 75.
Quality: Quality of youth engagement with program	Two items on the Facilitator Self-Assessment Tool regarding youth engagement were added together and the mean scores for each facilitator were calculated and the over-all mean for each treatment condition (*LN* and *RtR*) were calculated. Range in scores from 2 to 10.
Quality: Quality of youth engagement with program	Two questions on the *LN* and *RtR* Observer Assessment Tool were analyzed separately and together regarding youth listening and interacting during the training.
Quality: Quality of youth engagement with program	Six items on the Facilitator Alliance Scale were added together and mean scores for each Camp were calculated. The overall mean for each treatment condition (*LN* and *RtR*) were calculated. Range in scores from 6 to 30.
Quality: Quality of youth engagement with program	Nine items on the Group Cohesion Scale were added together and mean scores for each Camp were calculated. The over-all mean for each treatment condition (*LN* and *RtR*) were calculated. Range in scores from 9 to 40.
Program Differentiation: Between *LN*, *RtR* and *PoW*	Calculated the percent of items answered correctly for both the *LN* test and the *RtR* test and compared mean percentage scores for each condition (*LN vs. RtR vs. PoW*) using ANOVA. The goal was for participants in *LN* to score significantly higher on the *LN* test than for participants in *RtR* or *PoW*. The goal was for participants in *RtR* to score significantly higher on the *RtR* test than for participants in *LN* or *PoW*
Counterfactual: Experiences of counterfactual condition	All *PoW* data was calculated similarly to the intervention data.

Implementation element	Methods used to address each implementation element
Context: Other TPP programming available or offered to study participants (both intervention and counterfactual)	Percentages of youth who had ever had sex education in school were calculated for each group at baseline (63% of both *LN* and *RtR* youth and 58% of *PoW* youth had had sex education prior to coming to CHAMPS! Camp). During the course of the one year follow up, very few youth were exposed to additional sex education material in any group. There was no difference between groups
Context: External events affecting implementation	No events occurred that interfered with implementation.
Context: Substantial unplanned adaptation(s)	Percentage of missing activities

Notes: TPP = Teen Pregnancy Prevention

Appendix E: Randomized Cluster Analysis.

The multi-level RCT model, based on Bloom (2005) and the Cole, Deke & Zief Mathematica FAQs of 5/17/13, is:

Level 1: $Y_{ij} = \beta_{0j*} + \sum \beta_{kj*} X_{(k)ij}$

Level 2: $\beta_{0j} = Y_{00} + u_{0j} \beta_{kj} = Y_{k0} Y_{ij}$ = the outcome for individual i from cluster j (e.g. virgin/non-virgin; frequency of condom/birth control use; number of sexual partners; been or caused pregnancy, etc.). β_1 = the observed prevalence rates of the outcome (across the treatment and control conditions, = treatment indicator variable (e.g. 0 for Power of We, 1 for Reducing the Risk, or 1 for Love Notes). We conducted two analyses, one showing the impact for *RtR* relative to *PoW*, one showing the impact of *LN* relative to *PoW*). β_{0j} = Slope from level 2, based on the clusters in which individuals and treatments are nested. $\sum \beta_{kj*(k)ij}$ = Vector of covariates u_{0j} = random effect to adjust for the non-independence of individuals within clusters.

The multi-level hierarchical linear regression model was used to estimate program impacts for both continuous and dichotomous outcome measures. In the latter case, we also conducted analyses using the SPSS Logistic Regression models to check the robustness of the results both with and without the covariate. We employed cluster robust, heteroskedastic-consistent standard errors.

Appendix F: Sensitivity analyses using ANCOVA

To test whether the results presented in the report were sensitive to researcher decisions about how data were analyzed, we conducted a series of sensitivity analyses. The sensitivity analyses excluded variables that were included in the HLM analyses. The excluded variables included probability of being assigned to a cluster, the participant's cohort, the participant's standing on the baseline measure of the outcome variable, and the demographic variables of age and gender. Because the Equivalence Analysis determined that there were small but significant deviations from random distributions of ethnic groups across the 3 intervention conditions, the covariate consisting of the ratio of African-American and African participants to the sum of White, Hispanic and Asian participants was retained and employed throughout. Thus, the sensitivity analyses tested the impact of the treatment on the outcome variables with only the ethnicity covariate, and not all of the other control variables that had the potential to consume degrees of freedom and exclude cases with missing data.

Results of sensitivity analysis:

Comparison of the results of the HLM analyses reported above with the sensitivity analysis using ANCOVA below suggested the following:

1) The use of the Baseline outcome measures as covariates had minimal impact on results. An exception was the number of partners. There was a very low response rate for that variable during the pre-intervention assessment, so the use of that Baseline measure severely restricted sample size. The ANCOVA results are probably more meaningful. Because there was no significant effect, however, this is a moot point.
2) The use of the gender and age covariates had minimal impact on results.
3) The use of hierarchical linear modeling statistics to control for clustering had minimal impact on results.
4) A few specific effects differed across the analyses. Four outcomes were stronger in HLM:
    a. *RtR* had a significant effect on Ever Had Sex at the 3 month follow-up in the HLM analysis $(p < .03)$ but fell short of significance in the same ANCOVA analysis $(p < .08)$.
    b. *LN* had a significant effect on Ever Pregnant at the 6 month follow-up in the HLM $(p < .04)$ but that effect fell short of significance in the ANCOVA $(p < .09)$.
    c. *LN* had a significant effect on Ever Had Sex at the 6 month follow-up in the HLM $(p < .02)$ but that effect fell short of significance in the same analysis in the ANCOVA $(p < .10)$.

41

d. There was a marginal effect of *RtR* on Ever Had Sex at the 6 month follow-up in the HLM analysis $(p < .10)$ but that effect was nonsignificant in the ANCOVA $(p < .52)$.

Three outcomes were stronger in ANCOVA:

e. *RtR* had a significant effect on Had Sex in the Last 3 Months at the 3 month follow-up in the ANCOVA analysis $(p < .04)$ but fell short of significance in the same HLM analysis $(p < .06)$.

f. *RtR* had a significant effect on Sex Without Birth Control During the past 3 months at the 6 month follow-up in the ANCOVA analysis $(p < .001)$ but that effect fell short of significance in the HLM analysis $(p < .08)$.

g. *LN* had a significant impact on Had Sex in the Last 3 Months at the 12 month follow-up $(p < .01)$ but that effect was not significant in the HLM analysis $(p < .08)$.

Table F.1 Post-intervention ANCOVA estimated effects of Champs! treatments at 3 months.

Measure	Reducing the Risk *RtR*	Love Notes *LN*	Power of We *PoW*	*RtR* vs. *PoW* Mean dif	*RtR* vs. *PoW* p	*LN* vs. *PoW* Mean dif	*LN* vs. *PoW* P
Primary questions							
· Sex Without Condom Past 3 Months (%)	10.05	11.20	11.89	-.019	.428	-0.009	0.720
Sample Size	*378*	*357*	*328*				
· Sex Without Birth Control Past 3 Months (%)	7.14	8.64	11.89	-.047	.029	-.032	.143
Sample Size	*378*	*359*	*328*				
· Number of Partners Past 3 Months (mean)	.399	.523	.542	-.142	.225	-.012	.920
Sample Size	*303*	*304*	*291*				
Secondary Questions							
· Ever Pregnant (%)	1.33	1.95	2.74	-.014	.175	-.009	.404
Sample Size	*375*	*359*	*329*				
· Pregnant in Last 3 Months (%)	1.07	1.13	3.06	-.020	.044	-018	.063
Sample Size	*375*	*355*	*327*				
· Ever Had Sex (%)	30.34	34.08	36.47	-.062	.082	-.027	.459
Sample Size	*379*	*358*	*314*				
· Sex in Last 3 months (%)	18.47	21.01	24.92	-.065	.036	-.040	.204
Sample Size	*361*	*339*	*312*				

Notes: ANCOVA analysis used ethnicity as ac covariate, For number of partners, many participants reported "x", "-" or left the space blank; they were coded as missing. A $p \langle = .05$ is considered significant

Table F.2. Post-intervention ANCOVA estimated effects of Champs! treatments at 6 months.

Measure	Reducing the Risk *RtR*	Love Notes *LN*	Power of We *PoW*	*RtR* vs. *PoW* Mean dif	*RtR* vs. *PoW* p	*LN* vs. *PoW* Mean dif	*LN* vs. *PoW* P
Primary questions							
· Sex Without Condom Past 3 Months (%)	12.05	9.589	16.50	-.044	.093	-.068	.01
Sample Size	332	334	303				
· Sex Without Birth Control Past 3 Months (%)	9.06	8.04	17.22	-.081	.001	-.089	.0001
Sample Size	331	336	302				
· Number of Partners Past 3 Months (mean)	.37	.30	1.23	-.858	.031	-.907	.041
Sample Size	227	228	213				
Secondary Questions							
· Ever Pregnant (%)	2.42	1.49	3.64	-.012	.329	-.021	.091
Sample Size	331	335	302				
· Pregnant in Last 3 Months (%)	1.89	1.88	2.77	-.506	.613	-.441	.659
Sample Size	317	318	289				
· Ever Had Sex (%)	37.16	33.33	39.60	-.025	.517	-.064	.095
Sample Size	331	336	303				
· Sex in Last 3 months (%)	23.19	19.16	27.63	-.044	.188	-.084	.012
Sample Size	332	334	304				

Notes: ANCOVA analysis used ethnicity as ac covariate, For number of partners, many participants reported "x", "-" or left the space blank; they were coded as missing. A $p \langle = .05$ is considered significant

Table F.3 Post-intervention ANCOVA estimated effects of Champs! treatments at 12 months.

Measure	Reducing the Risk RtR	Love Notes LN	Power of We PoW	RtR vs. PoW Mean dif	RtR vs. PoW p	LN vs. PoW Mean dif	LN vs. PoW P
Primary questions							
· Sex Without Condom Past 3 Months (%)	13.20	13.33	14.24	-.012	.663	-.013	.643
Sample Size	356	344	309				
· Sex Without Birth Control Past 3 Months (%)	12.22	11.05	09.35	.027	.271	.012	.623
Sample Size	352	344	310				
· Number of Partners Past 3 Months (mean)	.63	.65	.69	-.054	.647	-.043	.725
Sample Size	278	265	237				
Secondary Questions							
· Ever Pregnant (%)	7.39	4.05	4.82	.026	.146	-.008	.674
Sample Size	352	346	311				
· Pregnant in Last 3 Months (%)	5.11	2.89	4.19	.010	.521	-.011	.473
Sample Size	352	346	310				
· Ever Had Sex (%)	37.16	33.33	39.60	-.025	.517	-.064	.095
Sample Size	331	336	303				
· Sex in Last 3 months (%)	23.19	19.16	27.63	-.044	.188	-.084	.012
Sample Size	332	334	304				

Notes: ANCOVA analysis used ethnicity as ac covariate, For number of partners, many participants reported "x", "-" or left the space blank; they were coded as missing. A $p \langle = .05$ is considered significant

Appendix G: Sensitivity analyses on Number of Partners using imputed data.

The primary impact analyses conducted on the participants' reported number of sexual partners in the past 3 months was based on a question that required them to enter a number, rather than endorse a fixed response. Some participants left the question blank, and some gave non-numerical responses such as "a few" or "many". Because we did not know if "a few", for example, meant three, five, seven or more partners, or if it meant the same such quantity to all participants, we did not code such responses. For the first of two sensitivity analyses on this question, however, we imputed zero partners for all participants who did not respond to this item, but who elsewhere reported that they either had never had sex, or had not had sex during the past three months. In the initial re-analysis, no numbers were imputed for participants who gave vague answers such as "few" or "many".

There was not a significant difference across conditions at baseline in the number of partners (F (2, 1309) = .463, p = .629. There appeared to be a conspicuous drop in number of partners from baseline to 3 months but it was not significant (F (1, 962) = .385, p = .535, nor was there a significant difference across training conditions at 3 months in the number of partners (F (2, 970) = 1.351, p = .259. The difference between training conditions in number of partners approached significance at 6 months (F (2, 970) = 1.351, p = .259, and was significant in the *RtR vs. PoW* conditions $(p \langle .043)$. This difference evaporated by 12 months F (2, 1061) = .432, p = .650.

Thus, the reanalysis produced one significant effect for *RtR* at 6 months in 8 contrasts.

Table G.1 ANCOVA estimated effects of Champs! treatments on Number of Partners During the Past 3 months with imputed data for virgins.

Measure	Reducing the Risk *RtR*	Love Notes *LN*	Power of We *PoW*	*RtR* vs. *PoW* Mean dif	*RtR* vs. *PoW* p	*LN* vs. *PoW* Mean dif	*LN* vs. *PoW* P
BL Number of Partners (mean)	.761	.811	.667	.094	.531	.144	.338
Sample Size	468	463	386				
3M Number of Partners (mean)	.227	.362	.319	.092	.292	.043	.625
Sample Size	343	325	302				
6M Number of Partners (mean)	.256	.194	.507	.313	.043	.252	.106
Sample Size	288	293	267				
12M Number of Partners (mean)	.363	.327	.387	.024	.715	.060	.361
Sample Size	372	368	321				

Notes: ANCOVA used baseline measure, age, gender, ethnicity, cohort and probability of assignment to cluster as covariates

A second reanalysis imputed a "0" for virgins and a "1" for participants who said that they had been sexually active during the past 3 months, but did not provide a number in response to the question about number of partners. In this reanalysis, there again was not a significant difference across conditions at baseline in the number of partners (F (2, 1335) = .439, p = .645, nor was there a significant difference across training conditions at 3 months in the number of partners (F (2, 1005) = .554, p = .575. The difference between training conditions in number of partners was significant at 6 months (F (2, 918) = 3.095, p = .046, and was significant in the *LN vs. PoW* conditions (p = .016), and approached significance in the *RtR* condition (p = .069). This difference was again gone by 12 months F (2, 1106) = .141, p = .868. Thus, the reanalysis produced two significant and one marginal effect at 6 months.

Table G.2 ANCOVA estimated effects of Champs! treatments on Number of Partners During the Past 3 months with imputed data for virgins (0) and sexually active nonvirgins (1) who did not report the number of partners.

Measure	Reducing the Risk RtR	Love Notes LN	Power of We PoW	RtR vs. PoW Mean dif	RtR vs. PoW p	LN vs. PoW Mean dif	LN vs. PoW P
BL Number of Partners (mean)	.768	.820	.683	.100	.247	.084	.337
Sample Size	487	480	396				
3M Number of Partners (mean)	.768	.820	.683	.057	.522	.032	.725
Sample Size	362	340	312				
6M Number of Partners (mean)	.406	.323	.666	.259	.069	.342	.016
Sample Size	317	318	292				
12M Number of Partners (mean)	.37	.377	.414	.037	.639	.037	.640
Sample Size	394	385	336				

Notes: ANCOVA used baseline assessment, age, gender, ethnicity, cohort and probability of assignment to cluster as covariates

With the consistent sample dataset, described in Appendix H, there again was not a significant repeated measures difference across training conditions in the number of partners (F (2, 805) = .515, p = .598, although there was a training by time interaction effect that approached significance (F (6, 2415) = 1.640, p = .132. Number of partners at 3 months was significantly lower than at baseline across conditions (M= .546 vs. .351, t(812) 23.149, p ⟨ .0001), but that number significantly increased at 6 months (M= .351 vs. .449, t(812) 20.104, p ⟨ .0001), only to fall again at 12 months (M = .449 vs. .334, t(812) 16.666, p ⟨ .0001). Because the sample was consistent, these fluctuations cannot be attributed to changing composition of the sample.

In the consistent sample, *RtR* was associated with a significantly lower number of partners that than *PoW* at both 3 months and 6 months (both $p \langle .0001$). Participants who were randomly assigned to *LN* happened to be significantly higher than those in the *PoW* control condition, which makes the finding that *LN* was significantly lower than *PoW* at 6 months was even more remarkable. Thus, this reanalysis produced significant reductions on number of sexual partners for *RtR* at 3 months and for both *RtR* and *LN* at 6 months.

Table G.3 Repeated measure ANCOVA estimated effects of Champs! treatments on Number of Partners During the Past 3 months with imputed data for virgins (0) and sexually active nonvirgins (1) who did not report the number of partners with consistent sample.

Measure	Reducing the Risk *RtR*	Love Notes *LN*	Power of We *PoW*	*RtR* vs. *PoW* t	*RtR* vs. *PoW* p	*LN* vs. *PoW* t	*LN* vs. *PoW* p
BL Number of Partners (mean)	.493	.656	.484	.255	.798	5.084	.0001
3M Number of Partners (mean)	.296	.370	.389	4.314	.0001	.886	.376
6M Number of Partners (mean)	.407	.326	.449	9.247	.0001	12.269	.0001
12M Number of Partners (mean)	.329	.344	.329	.017	.986	1.180	.239
Sample Size	280	276	257				

Note: The repeated measures ANCOVA used age, gender, ethnicity, cohort and probability of assignment to cluster as covariates

Appendix H: Sensitivity analyses on Ever Had Sex using a consistent composition sample.

Responses to the question "Have you ever had sexual intercourse" should remain stable or increase over time, and not decrease. When such scores appear to decrease, there are multiple potential causes: (a) A change in the composition of the sample from one time period to the next. If fewer non-virgins participate in later sessions, the proportion of respondents who honestly answered "no" would increase, causing the proportion to drop. (b) A change in the understanding of the question from one time period to the next. A participant could read "ever" to mean in one's lifetime on one occasion, and read "ever" to mean "recently" on a second occasion, causing the proportion to drop. (c) A change in reference group influencing what is seen by the participant as a socially desirable response. A participant who is a virgin might be embarrassed due to thinking about how peers might judge their behavior and deny at baseline, but be proud of it after 3months as a function of perceiving that is what the CHAMPS! staff and other program participants desire, causing the proportion nonvirgins to drop. Scores also could decrease (d) as a result of random error, due to careless responding.

To examine the impact of the consistency of the composition of the sample on responses to the virginity questions, repeated measures analyses were conducted on the question of "ever had sexual intercourse" and "had sexual intercourse in the past 3 months". The repeated measures analysis requires that each participant contribute responses at baseline, 3 months, 6 months and 12 months, plus have data entries for type of training, gender, age, ethnicity, cohort and probability of assignment to 2 or 3 study conditions. That led to the inclusion of n=711 participants and exclusion of n=716 cases for the first question, and inclusion of n=811 participants and exclusion of n=616 cases for the second question.

As Table H indicates, a drop in the report of Ever Had Sex from Baseline to 3 months was apparent in all three conditions in the reduced sample. This analysis indicated that differences in the composition of the sample was not the sole cause of the drop, leaving the possibilities of misunderstanding the question, social desirability responding or random error. The most likely cause is a combination of misreading the question and social desirability. There was a high correlation between responses to the "ever had sexual intercourse" question and the "had sexual intercourse in the last 3 months" at baseline (r = .671) 3 months (r = .674), 6 months (r = .672)

and 12 months (r = .679). It is possible that some participants confused the two points of reference.

Because the drop from Baseline to 3 months was most apparent in the *RtR* and *LN* groups, it seems likely that some participants in *RtR* and *LN* wanted to express their pride in their avoidance of sexual activity over the past three months, and not only made that declaration in response to the "had sexual intercourse in the last 3 months" item, but also in response to the "ever had sexual intercourse" item, even if meant contradicting their baseline report (which may have been accurate or inaccurate). Thus, responses to the "had sexual intercourse in the last 3 months" item is likely to have been more literally accurate, but both items reflect the tendency of *RtR* and *LN* participants to refrain from sexual activity.

The drop in rates shown in Appendix H is smaller than what is shown in Tables IV.1-IV.3. Therefore, the drop in the sexual initiation rates in Tables IV.1-IV.3 suggests that the composition of the samples may be somewhat different at each time point. Consequently, the Table IV.1-IV.3 partially reflect cross-sectional findings, with somewhat different analytic samples at each time point. The demonstration of baseline equivalence in Tables IV.1-IV.3, and the adjustment for background characteristics, helps to illustrate that each of these cross-sectional findings are credible, and this consistent sample analysis, although based on a small sample that may be unrepresentative because of the high levels of reliability in their participation, provides additional context for the longitudinal trends.

With the reduced but consistent sample and the "ever had sexual intercourse" question, the repeated measures ANOVA produced a significant effect of training (F (2, 708) = 7.132, $p \langle .001$), and a training by time interaction (F (6, 798) = 85.650, $p \langle .0001$), with the latter reflecting the greater differences between training and control conditions at 3 months and 6 months than at baseline or 12 months. There is no difference between *RtR* and *PoW* at baseline, but there is a significant difference between *RtR* and *PoW* at 3 months $\left(p \langle .0001 \right)$ and a difference that approaches significance at 6 months $\left(p \langle .067 \right)$. While there is a difference between *LN* and *PoW* at baseline, this does not threaten the validity of the HLM impact analyses, because baseline scores were always used as a covariate. In addition, the difference between *LN* and *PoW* at 3 months is significantly greater than the *LN-PoW* difference at baseline (t (466)

=2.494, $p < .01$), and the difference between *LN* and *PoW* at 6 months is also significantly greater than the *LN-PoW* difference at baseline (t (466) =3.887, $p < .0001$).

With the reduced sample and the "had sexual intercourse in the last 3 months" question, the repeated measures ANOVA produced a significant effect of training (F (2, 808) = 12.325, $p < .0001$), and a training by time interaction (F (4, 808) = 103.638, $p < .0001$), with the latter reflecting the greater differences between training and control conditions at 3 months and 6 months than at baseline or 12 months. Again, there is no difference between *RtR* and *PoW* at baseline, but there is a significant difference between *RtR* and *PoW* at 3 months $(p < .0001)$, 6 months $(p < .0001)$ and even 12 months $(p < .004)$. There is again a difference between *LN* and *PoW* at baseline, but the difference between *LN* and *PoW* at 3 months is marginally greater than the *LN-PoW* difference at baseline (t (530) = 1.778, $p < .076$), and the difference between *LN* and *PoW* at 6 months is also significantly greater than the *LN-PoW* difference at baseline (t (530) = 4.196, $p < .0001$). *LN* and *PoW* do not differ at 12 months.

Table H. Repeated measure ANCOVA estimated effects of Champs! treatments on Ever Had Sex and Had Sex in the Last 3 Months questions on consistent composition samples.

Measure	Reducing the Risk *RtR*	Love Notes *LN*	Power of We *PoW*	*RtR* vs. *PoW* T	*RtR* vs. *PoW* P	*LN* vs. *PoW* T	*LN* vs. *PoW* p
BL Ever Had Sex (%)	39.09	36.51	40.53	1.012	.312	2.791	.005
3M Ever Had Sex (%)	32.51	31.95	37.89	4.878	.0001	5.285	.001
6M Ever Had Sex (%)	37.45	31.54	39.65	1.839	.067	6.678	.001
12M Ever Had Sex (%)	37.04	38.59	37.89	1.008	.314	.808	.420
Sample Size	243	241	227				
BL Sex in Last 3 months (%)	23.30	21.04	24.61	1.378	.169	3.704	.0001
3M Sex in Last 3 months (%)	19.00	18.84	24.22	5.516	.0001	5.482	.0001
6M Sex in Last 3 months (%)	22.22	18.12	26.17	3.988	.0001	7.900	.0001
12M Sex in Last 3 months (%)	19.36	21.01	21.09	2.903	.004	0.127	.899
Sample Size	279	276	256				

Notes: Repeated measures ANCOVA used age, gender, ethnicity, cohort and probability of assignment to cluster as covariates

Appendix I: Sensitivity analyses on Ever Pregnant using a consistent composition sample.

Similar to the case with the question "Have you ever had sexual intercourse" responses to the question "Have you ever been pregnant or caused a pregnancy" generally should remain stable or increase over time, and not decrease. While there may be some individuals who believe at the 3 Month reporting period that they were pregnant or caused a pregnancy, and found out by the 6 month reporting period that was a false alarm, their numbers should be small.

As Table I.1 reveals, a drop in the pregnancy rate of the *LN* group from 3 months to 6 months was still apparent using the stable sample, so changes in sample composition could not explain the drop.

Table I.1 Repeated measure ANCOVA estimated effects of Champs! treatments on Ever Been Pregnant or Caused a Pregnancy questions on consistent composition sample.

Measure	Reducing The Risk *RtR*	Love Notes *LN*	Power of We *PoW*
BL Ever Pregnant (%)	00.00	00.00	00.00
3M Ever Pregnant (%)	1.25	2.50	2.21
6M Ever Pregnant (%)	2.08	0.83	3.10
12M Pregnant (%)	5.42	3.33	3.54
Sample Size	240	240	226

Notes: Repeated measures ANCOVA used age, gender, ethnicity, cohort and probability of assignment to cluster as covariates

To determine if data recording errors or social desirability bias might be causing the effect, we did a corrective recode of the 6 month reports. If a participant reported that they were pregnant or had causes a pregnancy at 3 months and provided that same answer at 12 months, their response as recoded to a "yes" (1) for 6 months. While that reduced the drop in the *LN* group, it did not completely eliminate it. The correction produced no changes in the *RtR* and *PoW* groups.

Table I.2 Repeated measure ANCOVA estimated effects of Champs! treatments on Ever Been Pregnant or Caused a Pregnancy questions on consistent composition sample with data correction at 6 months.

Measure	Reducing The Risk *RtR*	Love Notes *LN*	Power of We *PoW*
BL Ever Pregnant (%)	00.00	00.00	00.00
3M Ever Pregnant (%)	1.25	2.50	2.21
6M Ever Pregnant (%)	2.08	1.67	3.10
12M Pregnant (%)	5.42	3.33	3.54
Sample Size	240	240	226

Notes: Repeated measures ANCOVA used age, gender, ethnicity, cohort and probability of assignment to cluster as covariates

We also considered the possibility that a data recording error by participants at 3 months might be causing the apparent drop. We did an additional corrective recode of the 3 month reports. If a participant reported that they were never pregnant/had never caused a pregnancy at 6 months and provided that same answer at 12 months, their response as recoded to a "no" (0) for 3 months. As Table H.3 indicates, that produced a drop in pregnancy reports in all three groups, including a drop of 0.84% in *RtR*, 1.675% in *LN* and 1.32% in *PoW*. While we are not pleased by such errors, it should be noted that this amounts to recording mistakes by 2 cases in *RtR*, 4 cases in *LN* and 3 cases in *PoW*, or problems in 1.27% across the n=706 cases, whose average age was 15 years old.

Analysis of the corrected data indicated that pregnancies were lower in the *RtR* group than the *PoW* control group at both 3 months and 6 months $(p \langle .0001)$ and the *LN* group at 6 months $(p \langle .0001)$. There was an apparent rebound effect in the *RtR* group at 12 months, such that they reported significantly more pregnancies than the *PoW* control group. We will wait until the 24 month data to come in to insure that this is accurate.

Table I.3 Repeated measure ANCOVA estimated effects of Champs! treatments on Ever Been Pregnant or Caused a Pregnancy questions on consistent composition sample with data correction at 3 months and 6 months.

Measure	Reducing the Risk (*RtR*)	Love Notes (*LN*)	Power of We (*PoW*)	RrR vs. *PoW* t	RrR vs. *PoW* p	*LN* vs. *PoW* T	*LN* vs. *PoW* p
BL Ever Pregnant (%)	0.00	0.00	0.00				
3M Ever Pregnant (%)	0.41	0.83	0.89	7.107	.0001	.752	.452
6M Ever Pregnant (%)	2.08	1.67	3.10	5.179	.0001	7.394	.0001
12M Ever Pregnant (%)	5.42	3.33	3.54	5.986	.0001	.631	.529
Sample Size	240	240	226				

Notes: Repeated measures ANCOVA used age, gender, ethnicity, cohort and probability of assignment to cluster as covariates