Prevention Science (2019) 20:1074–1088
https://doi.org/10.1007/s11121-019-01032-2

Addressing the Adherence-Adaptation Debate: Lessons from the Replication of an Evidence-Based Sexual Health Program in School Settings

Jenita Parekh[1,2] · Elizabeth Stuart[3] · Robert Blum[1] · Valerie Caldas[2] · Brooke Whitfield[4] · Jacky M. Jennings[2]

Published online: 8 August 2019

Abstract
Whether high adherence to programs is necessary to achieve program outcomes is an area of great debate. The objectives of this study were to determine the frequency, type, and rationale of adaptations made in the implementation of an evidence-based program and to determine program outcomes for intervention program participants, as compared to comparison participants, by the level of adaptations. A total of 1608 participants in 45 classrooms participated. Percent adaptations was calculated by classroom. Thematic qualitative analysis was used to categorize types and rationales for adaptations. Program outcomes by level of adaptations were determined using logistic regression analyses and mean differences. Propensity score matching methods were used to create comparability between adaptation subgroup participants and comparison participants. Adaptations ranged from 2 to 97% across classrooms, with mean adaptations of 63%. Thematic analysis revealed that the adaptations made were related to delivery of content, rather than to the content itself and in response to participant needs and setting constraints. Program outcomes did not appear to be reduced for the high-adaptation subgroup. Understanding both rationale (intent) and type of adaptation made is crucial to understanding the complexity of adaptations. These finding support the argument for allowing facilitators some flexibility and autonomy to adapt the delivery of prescribed content to participant needs and setting constraints.

Keywords Program implementation · Evidence-based program · Sexual and reproductive health programs · Adolescents · Program implementation fidelity

Introduction

In the nation's largest coordinated effort in adolescent pregnancy prevention, the Family Youth and Services Bureau (FYSB) and the Office of Adolescent Health (OAH), provided over one billion dollars for the scale up and replication of adolescent sexual reproductive health evidence-based programs (SRH-EBPs) (Margolis and Roper 2014). This effort presents an opportunity for replication testing and investigation of implementation outcomes that may shed light on null or enhanced program effects (Fixsen et al. 2005; Nelson et al. 2012). Studies investigating implementation outcomes in evidence-based programs among SRH-EBPs and among adolescents are limited; most studies have been conducted among adults in the substance use, mental health, and physical health fields (Durlak and DuPre 2008; Metz et al. 2015).

Implementation refers to both how a program is delivered (e.g., adherence, adaptations) as well as how it is received (e.g., participant attendance). One implementation outcome is the "fidelity of implementation," which is defined as the extent to which facilitators execute a program as prescribed by design (Carroll et al. 2007a). Two important dimensions of implementation fidelity include adherence and adaptation (Dane and Schneider 1998; Durlak and DuPre 2008). Adherence is defined as

✉ Jenita Parekh
jenita.parekh@gmail.com

[1] The Department of Population, Family and Reproductive Health, The Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA

[2] The Johns Hopkins Center for Child & Community Health, Johns Hopkins Bayview Medical Center, 5200 Eastern Avenue, Suite 4200, Mason F. Lord Building, Center Tower, Baltimore, MD 21224, USA

[3] The Department of Mental Health, The Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA

[4] Child Trends, Inc., 7315 Wisconsin Ave, Suite 1200W, Bethesda, MD 20814, USA

faithfulness to the core content of a program (Berkel et al. 2011a; Carroll et al. 2007a; Dusenbury 2003). Adaptation is defined as any modification to a program, both modifications that make the program suitable for its context and partial completion of an activity. This differs from lack of fidelity, which is skipping or omitting an activity in its entirety (Blakely et al. 1987; Firpo-Triplett and Fuller 2012; Stirman et al. 2013). Adaptations may be made to content or content delivery (Stirman et al. 2013). Because implementation conditions in initial evidence-based settings cannot be replicated exactly in subsequent settings, studies that measure adaptations unfailingly report them during the replication of EBPs (Bumbarger and Perkins 2008; Castro et al. 2004; Durlak and DuPre 2008).

We chose to focus on adherence and adaptation because there is a debate in the field of implementation science regarding whether adherence is essential and whether adaptations negatively affect program outcomes (Cross and West 2011). Proponents of adherence suggest higher levels of adherence result in greater program outcomes (Battistich, Schaps, Watson, Solomon, & Lewis, n.d.; Blakely et al. 1987; Gottfredson et al. 1993; Kam et al. 2003), maintaining that adaptations are likely to be reactive and contribute negatively to program effectiveness (Bumbarger and Perkins 2008; Elliott and Mihalic 2004). Proponents of adaptations suggest that adaptations responding to local context are necessary and inevitable (Berman and McLaughlin 1978; Dusenbury 2003), and that adaptations suited to local context positively affect program outcomes (Barrera et al. 2017; Blakely et al. 1987; McGraw et al. 1996; Spoth et al. 2002). This debate may be in part due to variability in the definitions employed for adherence and adaptation and to poor, often retrospective, measurement of adherence and adaptation. Researchers and program developers are attempting to resolve some of the debate by blending the two views and considering adaptations appropriate if they do not interfere with core program components (Elliott and Mihalic 2004; Fixsen et al. 2005; McHugh et al. 2009).

Participant responsiveness, which includes attendance, is also an important implementation outcome; participants must be present and engage or participate in the intervention for the intervention to produce effects (Carroll et al. 2007b; Century et al. 2008). Adherence and attendance are conceptualized to have a bidirectional relationship; attendance has been conceptualized as both a moderator of adherence and program outcomes (Berkel et al. 2011b; Carroll et al. 2007b), and as a mediator between pre- and post-program outcomes, and adaptations and program outcomes (Kelsey and Layzer 2014a; Schoenfelder et al. 2013). For example, the relationship between adherence and outcomes may be moderated by attendance because if participants do not attend the intervention or receive it well (regardless of adherence), there will not be an effect on outcomes. Consequently, adaptations that take into account participants' needs are expected to increase participant attendance, responsiveness, and program outcomes.

There have been several limitations in prior studies examining adaptations and adherence. Prior studies have calculated frequencies of adherence and adaptations, listed types of adaptations, or linked adherence/adaptation scores to program outcomes, but have rarely conducted all three (Ennett et al. 2011; Kershner et al. 2014; McGraw et al. 1996; McHugh et al. 2009; Moore et al. 2013; Wang et al. 2015). There are many types of adaptations. Without understanding the type of adaptation, conclusions or hypotheses about adaptations are made without context. The objective of the present study was to determine: (1) the frequency of adherence and adaptations in the implementation of an SRB-EBP, *Making Proud Choices*; (2) the type and rationale for adaptations made; and (3) program outcomes for intervention and comparison participants, by the level of adaptations made in the classroom (high/middle/low adaptation). This study employed an innovative statistical approach to utilize both implementation and control participants, and an in-depth mixed methods approach, to explore adherence and adaptations, so that causal inferences can be made regarding program outcomes. Because of previous research and theory support associations between attendance and program outcomes, a sensitivity analysis was conducted to ensure the program outcomes observed by adaptation subgroup were not affected by attendance.

Methods

Study Overview

The present implementation study was part of a larger quasi-experimental parent evaluation study conducted from 2013 to 2015 in communities where at least half of program participants resided in communities of high teen pregnancy and STI prevalence. One program site from the parent study was selected to participate in this implementation study, *Making Proud Choices,* based on minimal missing entries, a robust sample size to detect effects, variability in implementation outcome measures, and fewer than 5% of activities omitted (to ensure the relationship between adaptations and program outcomes was not confounded by content omission). The Johns Hopkins School of Medicine IRB approved all study protocols.

Facilitators implemented the program with middle school students in 46 school-based classrooms across 14

schools during normal school hours. Program facilitators were personnel from outside the school. Depending on the time of the year and staff turnover, there were up to three facilitators total. One facilitator was assigned to each classroom and taught multiple classrooms. The evaluation design included matched intervention and comparison groups at the classroom level. The intervention group was first recruited, then the comparison group was recruited based on similarity in geographic location and grade. In some cases, the comparison group was recruited from the same school, if allowed. In instances where school officials did not want to split up a grade, comparison participants were recruited from the grade below. In some cases, comparison participants were recruited from another school in the same town. Comparison group participants received information on a non-sexual topic.

Evidence-Based Program

Making Proud Choices is a comprehensive HIV/STI/pregnancy prevention program, consisting of eight 1-hour modules of group discussion, role play, videos, and exercises. The program addresses goals, consequences of STIs/pregnancy, beliefs around HIV/STIs and condom use, prevention strategies, and negotiation skills. The developer recommends that the entire intervention be completed within 2 weeks in school or community-based settings.

Data Collection

Study staff collected demographic and outcome data by administering surveys to participants at pre- and post-test. Staff administered pre-test surveys on the first day of the intervention and post-test surveys on the last day, approximately 4 to 6 weeks after pre-test. Staff entered survey, adherence, and adaptation data into an online data collection system on a quarterly basis. Program staff collected participant attendance data on an attendance log.

Measures

Implementation Measures

The program developer created the fidelity log. The facilitator who delivered the session completed the log, for every classroom, within 48 h of the session. The fidelity log was comprised of 34 required *Making Proud Choices* activities, for which a facilitator selected: (1) "completely finished activity," (2) "finished activity with changes," or 3) "did not finish activity." When a program facilitator marked "finished activity with changes," the facilitator was prompted to describe and explain the rationale for changes made. To limit social desirability bias, a training on how to complete fidelity logs was conducted emphasizing honest reporting. In the training, program facilitators were instructed to mark deviations from the prescribed program as a "change" and to describe this in detail. Facilitators received between 16 and 24 h of training on program content and implementation strategies, dependent on the facilitator's level of knowledge and experience. Although facilitators received training on how to fill out the fidelity logs, they were not provided training on how to make adaptions.

For this study, adherence is measured as a percent of the number of times the program facilitator selected the box "completely finished" out of the total number of activities offered ($n = 34$). Adaptation is measured as a percent of the number of times the program facilitator selected "finished activity with changes" out of the total number of activities offered ($n = 34$). Non-completion or omission is measured as a percent of the number of times the program facilitator selected "did not finish activity" out of the total number of activities offered ($n = 34$). For the purposes of this study, non-completion or omission is not considered an adaptation, rather it is considered lack of fidelity. Attendance is measured as a percent of the number of sessions an individual attended out of the total number of sessions delivered. The number of sessions delivered differed by classroom and implementation setting. On average, classrooms completed the intervention in 4 to 6 weeks. This variation in length was due to school constraints such as delayed programming for holidays or inclement weather. Additionally, some schools allowed only 45 min for implementation while others allowed 90 min.

Program Outcome Measures

The evaluation used validated pre- and post-test surveys from the Personal Responsibility Education Program study. These self-report measures were assessed for reliability and validity and were reviewed by a panel of experts (Margolis and Roper 2014). The two primary outcomes of interest were: sexual reproductive health knowledge and behavioral intent. Sexual reproductive health knowledge was measured at pre- and post-test. A refined scale of nine items was used to create a composite score of knowledge. Questions included five true/false items about behaviors that put one at risk for contracting HIV and four items regarding effective methods to protect people from STIs and pregnancy (e.g., douching). A composite knowledge score was created for each individual as the percentage of items answered correctly out of the total number of items. Higher scores represented greater SRH

knowledge. Behavioral intent questions were measured at post-test for intervention and comparison groups and used a five-point Likert scale (much more likely to much less likely) to gauge program impact. Questions measuring behavioral intent were: (i) likely to use or ask a partner to use birth control in the next 6 months, (ii) likely to use or ask a partner to use condoms in the next 6 months, and (iii) likely to abstain from sexual intercourse in the next 6 months. These variables were dichotomized in analyses to allow for easier interpretation with a logistic regression.

Independent Variables and Covariates

Demographics assessed include: (i) age, (ii) ethnicity, (iii) race, (iv) gender, and (v) mother's education. Baseline risky sexual behavior includes: ever had sex, ever pregnant, sex in the past 3 months, SRH knowledge composite score at pre-test, condom use in the past 3 months and birth control use in the past 3 months (all the time, most of the time, some of the time, none of the time). Social emotional competence measures include: resisted or said no to peer pressure, cared about doing well in school, shared things that matter with a parent/guardian, and managed conflict without causing more conflict (all of the time to none of the time). See Appendix Table 3 for more information on measures.

Data Analysis

Study Design

This study is a mixed methods evaluation that employed a convergent design. As described by Fetters, this design is when qualitative and quantitative data are collected and analyzed during a similar timeframe and then systematically integrated into the findings (Fetters et al. 2013). The author of this study first collected quantitative data from fidelity logs on percent adherence and percent adaptations (Obj 1). Then the author collected qualitative data from fidelity logs using open-ended facilitator responses on the type and rationale for adaptions made during programming (Obj 2). To integrate these data, the author converted the qualitative data into quantitative data by coding frequencies of responses. This mixed methods design allowed the author to integrate the qualitative and quantitative data and provide additional information on why facilitators may not have completed all program activities or completed them with changes. This data was further used to explain the program outcomes observed (Obj 3). All statistical analyses were conducted using STATA 12.0, R (MatchIt) software, and Atlasti (ATLASti. Version 1.0.24 1999). Data analysis procedures are described by objective.

Objective 1. Using fidelity log entries, frequency calculations were conducted for three categories: adherence to core content, adaptation, and non-completion of core content.

Objective 2. Fidelity logs provided free text space to indicate the type and reason for adaptations, which were coded and analyzed for emerging categories (for both type and rationale of adaptation). The qualitative codes were then converted into frequencies.

Objective 3. Scale (item analysis) and reliability testing were conducted on two SRH knowledge scales (total of 11 items) related to pregnancy, HIV, and STI acquisition and protection. Item analysis indicated that participants answered two of the items incorrectly at a much higher percentage than other items and were dropped from the scale. Cronbach's alpha was used to estimate reliability (Cronbach 1951). The resulting scale of nine items had a reliability of .723, which is considered sufficient internal consistency.

To determine if adaptations had an effect on program outcomes, three subgroups of classrooms were created: low adaptation (adaptation < 25%), middle adaptation (adaptation between 25 and 50%), and high adaptation (adaptation > 50%). There is limited literature on adaptation levels and positive results, however, Durlak and Dupre's review of implementation literature reports positive results with fidelity levels between 60 and80% (Durlak and DuPre 2008), so the authors set 75%+ adherence as a cut off for the low adaptation group. To obtain cutoffs for the other subgroups, the histogram and sample size was taken into consideration to create equal groups at natural cutoffs.

High participant attendance is statistically associated with greater participant program outcomes (Blake et al. 2001). Given the link between attendance and program outcomes and the difference in mean attendance in each of the three subgroups, two analyses were conducted to ensure program outcomes by adaptation subgroup were not affected by attendance: (1) primary analysis in which participants with less than 75% session attendance or missing attendance (< 10% of participants) were excluded and (2) a sensitivity analysis including all participants. The primary sample included 535 intervention and 757 comparison participants. The sensitivity analysis sample included 851 intervention and 757 comparison participants. Only participants who completed the post-test were

included in analyses of post-test outcomes. Due to attrition, the post-test sample included 493 intervention and 558 comparison participants (1051 total) for the primary analysis and 689 intervention and 558 comparison participants (1247 total) for the sensitivity analysis (see Fig. 1).

Propensity score matching methods were used to create comparability between adaptation subgroup participants and comparison participants (Rosenbaum and Rubin 1983; Stuart and Green 2008). Observed covariates used for matching included: all demographic variables, baseline risky sexual behavior variables, and baseline social emotional competence because these covariates have demonstrated links to program outcomes in the literature (Mmari et al. 2005; Wamoyi and Wight 2014). Full matching was used to match treatment group members to comparison group members with similar propensity scores for each adaptation subgroup. Full matching makes use of all individuals in the data and has been shown to be effective at reducing bias due to observed confounding variables (Stuart and Green 2008). Differences in mean knowledge score were calculated in addition to the outcome model.

To estimate the program outcomes for each subgroup, the following steps were conducted: (1) Fitted a propensity score model of an adaptation subgroup (low/medium/high adaptation) as a function of covariates using the adaptation subgroup from the intervention group and the full comparison group. (2) Used full matching to match treatment group members to comparison group members with similar propensity scores (Hsieh and Shannon 2005). (3) Checked the balance of covariates in the matched intervention and comparison groups and controlled for imbalances (> .20 diff) in the outcome model if necessary. (4) Fitted a model of the outcome as a function of adaptation subgroup and covariates with the full matching weights. (5) Repeated steps 1–4 with remaining two adaptation subgroups.

Less than 0.1% of fidelity data were missing and for these values, a value of "completed fully" was entered. Less than 6% of the outcome variables, among those who answered the post-test, had missing data. Only members for which outcome variables were observed were used in analysis (Sterne et al. 2009). Covariates are required to have a non-missing value in R software prior to propensity score matching. Missingness for covariates ranged from 1 to 22%. For covariates with missing values, the mean score of the covariate was imputed prior to propensity score matching. For variables with > 10% missingness, a dummy variable was created and included in the propensity score model so that patterns of missingness were matched in the propensity score analyses. Because participants were nested within classrooms, clustering was accounted for at the classroom level. In instances where 1–3 items in the scale were missing for the composite SRH knowledge score, the observed items were used to create the score. If a participant responded to three or less items in the scale, the composite SRH knowledge score was coded as missing.

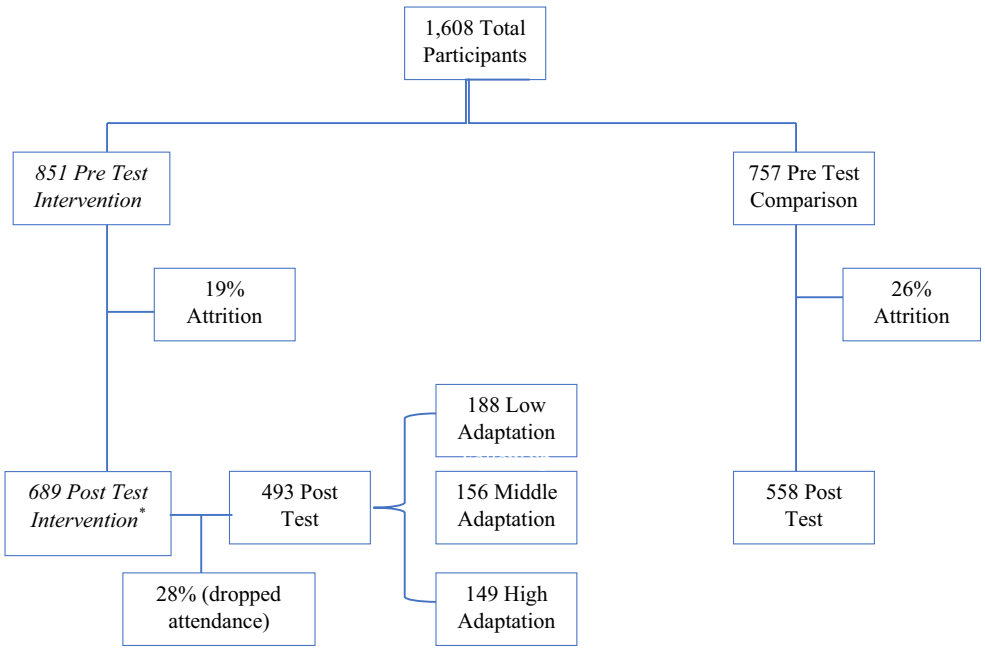

Fig. 1 Sample size and attrition for NJPREP implementation of making proud choices from 2013 to 2015, $N = 1051$

*Denotes intervention sample used for sensitivity analysis.

Results

Study Population

The high attendance (≥ 75%) intervention and comparison participants can be stratified into low-, medium-, and high-adaptation subgroups at post-test. High attenders did not portray significant differences in baseline sexual risk behaviors as compared to those who had missing or low attendance in the intervention group, differing significantly only on report of mother's education (47% versus 39% for "greater than a GED/HS education"). Participants in both the intervention and comparison group who answered the pre-test but did not answer the post-test had significantly greater percent distribution of males (58% vs. 49%), greater percent of African Americans (59% vs 43%), greater percentage of participants who had ever had sex (29% vs 21%), and a greater percentage of participants who used a condom at last sex (21% vs 16%). Participants were, on average, 14.8 years (SD = .06), 51% female, 45% African American, 17% Caucasian, and 40% Hispanic. At baseline, roughly 19% of the participants had ever had sex, with 15% of total participants having sex in the past 3 months. See Table 1.

Table 1 demonstrates significant differences in key baseline covariates between each of the adaptation subgroups and the comparison group, pre-matching. Significant differences in baseline covariates are indicated by *. Propensity score matching resulted in no significant differences in baseline covariates between the adaptation subgroup and comparison group.

Frequency of Adherence and Adaptation to Program Core Components

Percent adaptations to core content in classrooms ranged from 3 to 98%, with a mean adaptation of 63%. Mean adherence was 37%. Non-completion was below 6% across classes.

Type and Rationale for Adaptations Made

Most adaptations made (53%) were related to an increase in dosage, both in the number of sessions and in overall time to complete the program. On average, it took facilitators 11 h to complete the 8-h program. One reason for this is there was not enough time to deliver program content in the allotted time prescribed by the program. There were additional behavioral issues that required stopping the class and school constraints that took away from time to teach (e.g., unannounced field trips, inclement weather, state-mandated student testing, school-wide drills). The second and third most common adaptations involved changing the way the activity was presented to the group (15%). For example, facilitators explored a topic through discussion instead of playing a game, and partially completing the content (7%). These were also primarily due to time constraints and/or participant behavior. Approximately 5% of adaptations were related to translation of modules so that Spanish speaking students could understand the lessons, another 5% involved adding contraceptive models to the lesson, and 8% were "other" adaptions (e.g., change in activity sequence).

Comparing Program Outcomes for Low-, Middle-, and High-Adaptation Groups

There were significant differences in mean attendance for the three adaptation subgroups [F = 9.57, $p = 0.001$]. Post hoc comparisons using the Tukey HSD Test showed that the mean attendance score for the high-adaptation subgroup (M = .80) was significantly different than the middle (M = .84) and low adaptation subgroup (M = .87), however, there were no significant differences between the middle- or low-adaptation subgroup. Table 2 presents study results, selecting for high attendance. Program outcomes did not appear to be reduced for the high-adaptation subgroup. Program outcomes comparing the intervention condition to the comparison condition for the low-, middle-, and high-adaptation groups show that differences in SRH knowledge scores were greatest for the middle- and high-adaption groups. Intentions to abstain from sex showed similar intervention versus comparison odds across groups [low: OR = 1.63, $p = .17$; middle: OR = 1.43, $p = .23$; high: OR = 1.34, $p = .37$]. The high adaption group also showed higher odds of intervention participants intending to use birth control in the next 6 months compared to the comparison group [low: OR = 2.29, $p = .01$; middle: OR = 2.36, $p = .01$; high: OR = 5.67, $p = .00$] and higher odds of intervention participants intending to use condoms in the next 6 months [low: OR = 2.04, $p = .04$; middle: OR = 1.96, $p = .04$; high: OR = 3.10, $p = .04$].

Sensitivity Analyses

Because previous research and theory support associations between attendance and program outcomes, a sensitivity analysis was conducted to ensure that the program outcomes by adaptation subgroup were not due to attendance but due to adaptations. These analyses included all participants, regardless of attendance status. Program outcomes comparing all intervention participants to comparison participants for the low-, middle-, and high-adaptation groups

Table 1 Baseline characteristics for participants in low-, middle-, and high-adaptation groups at post-test who attended 75% of the *Making Proud Choices* sessions, $N = 1051$

	Low adaptation		Middle adaptation		High adaptation		Comparison		Total	
	n	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)
Demographics										
Age	188	15.2 (0.15)	156	14.6 (0.17)	149	14.3 (0.18)*	558	14.9 (0.08)	1051	14.8 (0.06)
Gender										
Male	95	50	81	52	65	44	271	49	515	49
Race										
African American	97	51*	80	51	66	44	227	41	471	45
Caucasian	29	15*	27	17	26	17	97	17	181	17
Asian	10	5*	14	9	7	5	66	12	97	9
Native American	14	7*	9	6	12	8	31	6	66	6
Hawaiian	9	5*	3	2	13	9	33	6	58	6
Ethnicity										
Hispanic	75	40	50	32	76	51*	213	38	416	40
Mother's education										
Less than HS	29	15*	20	13*	19	13*	75	13	144	14
HS/GED	54	29*	36	23*	49	33*	123	22	262	25
More than HS	85	45*	88	56*	62	41*	237	42	473	45
Sexual knowledge and behavior										
SRH knowledge	188	77 (0.20)	156	70 (0.02)	149	69 (0.02)*	558	73 (0.01)	1051	72 (0.01)
Sex, ever	54	29*	31	20*	21	14*	92	17	198	19
Pregnancy, ever	2	1	0	0	3	2*	3	1	5	1
Sex, past 3 months	44	23*	25	16	19	13*	68	12	156	15
Birth control past 3 months, always	14	7*	14	9	10	7	29	5	67	6
Condom use past 3 months, always	21	11*	14	9	11	7	38	7	84	8
Social emotional competence										
Cared about doing well in school, always	107	57*	82	53*	85	57*	246	44	522	49
Said no to peer pressure, always	63	33*	38	24*	44	30	166	30	312	30
Shared with parent, always	28	15	31	20	36	24*	89	16	184	17
Managed conflict, always	45	24*	20	13*	27	18	95	17	187	18

*Significantly different ($p <= .05$) from comparison group pre matching. No significant differences post matching

show similar findings, indicating that program outcomes by adaptation subgroup were not due to attendance.

Discussion

Results indicate that there is variability in frequency of adaptations made and percent adaptations made, by classroom, in the implementation of the program. The type of adaptations suggest that program implementers were not making changes to core content of the program, but rather were making changes in delivery of the content such as dosage, modality, and adding props, as a response to classroom context and participant needs/behavior. These data suggest that there is high completion of core content, but not without changes to the delivery of content. These data also suggest that the program facilitators were aware of the importance to adhere to dosage/pedagogy and marked any deviation from these core components as an adaptation. This is likely due to the reiteration of the evaluation team to note any deviation from the prescribed program on the fidelity log.

As compared to the comparison condition, level of adaptation (low, middle, and high) produces similar program outcomes in knowledge, intent to abstain, and intent to use birth control; and the high-adaptation subgroup does not appear to negatively affect program outcomes. This is not surprising, given the nature and type of adaptations

Table 2 Program outcomes for low-, middle-, and high-adaptation groups among high attenders in NJPREP, *Making Proud Choices*, 2013–2015

	Low-adaptation adherence > = 75%			Middle-Adaptation adherence 50–75%			High-Adaptation Adherence < = 50%		
	Mean adherence 84.41% Mean adaptation 15.59%			Mean adherence 65.64% Mean adaptation 34.12%			Mean adherence 37.6% Mean adaptation 62.0%		
	Intervention *n* 188	Comparison *n* 558	% Diff	Intervention *n* 156	Comparison *n* 558	% Diff	Intervention *n* 149	Comparison *n* 558	% Diff
Knowledge % correct	89.7%	75.4%	14.3%*	86.6%	69.2%	17.4%*	86.9%	69.1%	17.8%*
	OR	CI	*p* value	OR	CI	p value	OR	CI	p value
Intent to abstain from sex in next 6 months	1.63	0.80–3.30	0.17	1.43	0.79–2.61	0.23	1.34	0.69–2.63	0.37
Among those who do not plan to abstain									
Intent to use birth control (including condoms) in next 6 months	2.29[a]	1.28–4.09	0.01*	2.36[b]	1.09-4.13	0.01*	5.67[c]	2.51-12.85	0.00*
Intent to use condoms in next 6 months	2.04[a]	1.11–3.76	0.04*	1.96[b]	1.05-3.66	0.04*	3.10[c]	1.01-9.51	0.04*

*Significant differences within adaptation groups between the intervention and control participants at the 5% level

[a] Intervention *n* 101; Comparison *n* 298

[b] Intervention *n* 102; Comparison *n* 298

[c] Intervention *n* 89; Comparison *n* 298

being made. The adaptations made were largely a reaction to time constraints and student environment.

This paper presents an analysis on high attenders (participants who attended 75% or more of sessions). This subsample of high attenders was used for analysis due to the known linkage of program attendance and program outcomes. While selecting for high attenders reduces the intervention sample size, a sensitivity analysis including all participants was conducted to show there were no differential rates of attrition and to examine program impacts on the full sample. The high-adaptation subgroup produced significant intent to use a condom compared to the comparison condition, while the middle- and low-adaptation subgroups do not. This is likely because of the significant differences in attendance between subgroups. Even so, this is in line with the hypothesis that intervention participants in the high-adaptation subgroup do not appear to have reduced program outcomes. An argument in the larger debate is that planned adaptations are appropriate and unplanned adaptations are inappropriate. However, as evidenced by these data, not all adaptations were planned; for example, student behavioral issues. It is important to note that the implementation team was able to schedule extra time to complete the program with partner sites; however, this may not always be the case. Planning for more sessions than the required program dosage may be an appropriate planned adaptation for future implementers.

Given that adaptations do and will occur, our findings support researchers and program developers that consider adaptations appropriate if they do not interfere with the core components of a program (Elliott and Mihalic 2004; Fixsen et al. 2005; McHugh et al. 2009). Additionally, our findings support the argument for allowing facilitators some flexibility and autonomy to adapt the delivery of content of EBPs to participant needs and setting constraints. Developers themselves are not entirely clear on what activities are necessary and what could be omitted under time constraints (Philliber 2015), therefore additional studies should be conducted to test core curriculum components and assess what program information can be adapted and what needs to be adhered to. EBP trainings typically do not include extensive training on how to make adaptations or proactively plan for them, and implementers and investigators have noted that they "struggled with the process of adaptation" (Bell et al. 2007; Kelsey and Layzer 2014b). Program implementers are often instructed to contact the program developer regarding adaptations they would like to make, rather than being taught self-efficacy to adapt the program (Ebbole 2007). This is particularly problematic for implementers who are often making adaptations in real time to react to unpredictable contexts and behaviors in the classroom. Incorporating sessions on how to make adaptations, plan for increasing dosage, or handle

instances of time constrains, and behavior issues might be useful for facilitators.

Limitations and Strengths

This study was conducted within a funding environment of the Family and Youth Services Bureau, where fidelity was encouraged. It was also implemented in a school-based setting, so generalization beyond that setting is not warranted. Types of adaptations may differ in other contexts. The literature is mixed as to whether self-reported fidelity data are correlated with observational measures (Cross and West 2011; McGraw et al. 1996); however, as described in the methodology, steps were taken to limit social desirability bias. Similar studies of sexual reproductive health programs typically confirm they maintain "high fidelity," but often do not report percentages, apart from a couple studies (Feldman Farb and Margolis 2016; Kershner et al. 2014). This study does not suggest that adaptations appear to affect program outcomes; rather that adaptations of this particular nature and frequency listed in this study do not appear to reduce program outcomes. Additionally, the literature discusses the need to measure fidelity and adaptions as separate constructs (as opposed to "completely finished," "finished activity with changes," and "did not finish") (Blakely et al. 1987; Berkel et al. 2011a, b), however, the author was limited to working with the available program tools designed for the curriculum. Future studies may create additional categories to more accurately tease out fidelity and adaptations (fully/partially/no completion, with prompts for adaptations in each of those categories).

Prior studies (Kershner et al. 2014; McGraw et al. 1996; Moore et al. 2013) examining adaptations and fidelity calculate frequencies of fidelity and adaptations, list type of adaptations, or link fidelity/adaptation scores to program outcomes. This study utilizes a mixed methods approach to address all three aspects of prior studies. Because implementation data are often not collected for comparison groups, analyses linking implementation to program outcomes are purely descriptive. This study employed propensity score matching to utilize both implementation and comparison participants, so that causal inferences can be made regarding program outcomes. One limitation is that pre-test data were not collected on behavioral intent questions. This was due to the nature of these questions asking about participants' changes in intentions due to program participation. However, we feel confident that since we used OAH-approved and -validated surveys, and matched participants on 15 covariates, including demographics, sexual knowledge and behaviors, and social emotional competence, these findings have validity. Additionally, this study is focused on looking at adaptation and program outcomes, rather than actual change in sexual behavior, therefore, inability to match participants on behavioral intent should not hinder the study.

This study found that adaptations were often made to include more skill-building opportunities or to increase the amount of time spent implementing the program to ensure complete program implementation. However, these adaptations are not necessarily typical to adaptations made across schools. Adaptations to school interventions outside a study setting may be more likely to make significant cuts and/or drop activities they do not like or cannot facilitate. Without a comparison group of high adapters who were allowed autonomy to make revisions as they saw fit, such as cutting out lessons, this study is unable to determine whether such flexibility would have led to positive or negative outcomes. However, this study does find that focused adaptations that enhance existing lessons, such as the addition of skill-based activities and increased time, can maintain program effectiveness.

Conclusion

These findings support a flexible, blended view of adherence and adaptation, and support a movement away from the delineation of adaptations as "good" or "bad," and instead take into consideration their complexity. Measuring both rational and type of adaptation made is crucial to understanding the complexity of adaptations and their influence on program outcomes.

Funding Its contents are solely the responsibility of the author and do not necessarily represent the official views of the funder. J. Parekh was supported by the National Institute of Allergy and Infectious Disease (NIAID T32 AI050056-12).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of Johns Hopkins University's IRB and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Funding Information This project is supported by the Administration for Children and Families' Family and Youth Services Bureau (FYSB), utilizing federal funds HHS-2010-ACF-ACYF-PREP-0125.

Appendix

Table 3 Measurement variable table indicating question, response categories, operationalization, and source

Variable	Question	Response categories	Operationalization	Source
Dependent Variables — measured for Intervention and comparison participants				
Knowledge of risky behaviors scale	Can the following behaviors put you at risk for HIV? a) Sharing needles for tattooing and piercing b) Having vaginal sex without a condom c) Donating blood* d) Using the same condom twice e) Hugging f) Having anal sex without a condom Which of the following methods are effective if used correctly to protect people from STDs (including HIV) and pregnancy? a) choosing not to have sex (abstinence) b) Using hormone based birth control (e.g., the pill, Depo-Provera shot, patch, vaginal ring) c) Using condoms d) using withdrawal* e) douching (washing out the vagina) * denotes items that were dropped in refined scale	Yes/no/not sure	Continuous (mean scale score)	PREP post-survey
Knowledge of methods for prevention of pregnancy and STD scale	Would you say that being in the program has made you more likely, about the same, or less likely use or ask a partner to use birth control in next 6 months	Protects from pregnancy & STD/HIV/protects from pregnancy only/protects from neither		
Behavioral intent	Would you say that being in the program has made you more likely, about the same, or less likely to use or ask a partner to use condoms in next 6 months	Much more likely/somewhat more likely/about the same/somewh at less likely/much less likely/I will abstain from sexual intercourse (choose not to have sex) in the next 6 months	Dichotomized: 1: Much more likely/somewhat more likely 0: About the same/somewhat less likely/much less likely	PREP post and 3-month follow-up survey
Behavioral intent	Would you say that being in the program has made you more likely, about the same, or less likely to use or ask a partner to use condoms in the next 6 months likely to abstain from sexual intercourse in the next 6 months	Much more likely/somewhat more likely/about the same/somewhat less likely/much less likely/I will abstain from sexual intercourse (choose not to have sex) in the next 6 months	Dichotomized: 1: Much more likely/somewhat more likely 0: About the same/somewhat less likely/much less likely	PREP post- and 3-month follow-up survey
Behavioral intent	Would you say that being in the program has made you more likely, about the same, or less likely to use or ask a partner to use condoms in next 6 months likely to abstain from sexual intercourse in next 6 months	Much more likely/somewhat more likely/about the same/somewhat less likely/much less likely	Dichotomize d: 1: Much more likely/somewhat more likely 0: About the same/somewhat less likely/much less likely	PREP post-and 3-month follow-up survey
Behavior	Did you have sex in the past 3 months?	Yes/no	Dichotomous	3-month follow-up survey
Behavior	In the past 3 months, with how many people did you have sexual intercourse, even if only one time	1 person, 2–3 people, or 4 or more people	Dichotomized 1: 2+ 0: 1 person	3-month follow-up survey
Behavior		All of the time/most of the time/some of the time/none of the time	Dichotomized 1: All of the time	3-month follow-up survey

Table 3 (continued)

Variable	Question	Response categories	Operationalization	Source
	When you had sexual intercourse in the past 3 months, how often did you or a partner use birth control?		0: Most, some, none	
Behavior	The last time you had sex, did you or your partner use a condom?	Yes/no/do not remember	Dichotomized "Do not remember" coded as missing	3-month follow-up survey
Implementation variables — measured for intervention participants only				
Attendance	Did you feel interested in program sessions and classes? Did you feel the material presented was clear? Did discussions or activities help you to learn program lessons? Did you feel respected as a person? Were you picked on, teased, or bullied in this program? Did you have a chance to ask questions about topics or issues that came up in the program? Were youth in this program picked on, teased, or bullied because people thought they were lesbian, gay, bisexual, or transgender? Were youth in the program picked on, teased, or bullied because of their race or ethnic background?		Continuous (mean scale score)	
Participant responsiveness	Did you feel interested in program sessions and classes? Did you feel the material presented was clear? Did discussions or activities help you to learn program lessons? Did you feel respected as a person? Were you picked on, teased, or bullied in this program? Did you have a chance to ask questions about topics or issues that came up in the program? Were youth in this program picked on, teased, or bullied because people thought they were lesbian, gay, bisexual, or transgender? Were youth in the program picked on, teased, or bullied because of their race or ethnic background?	Five-point Likert All of the time/most of the time/some of the time/a little of the time/none of the time	Continuous (mean scale score)	
Fidelity subdomain: core content	For each activity, indicate whether the activity was fully completed, completed with changes, or not completed. Place a check in the appropriate column	Completed/completed with changes/did not complete	Adherence = % core activities completed/total core activities Adaptation = % core activities completed with changes/total core activities	Facilitator fidelity logs
Independent variables — measured for intervention and comparison participants				
Intervention receipt		A value of "0" is assigned to participants in the comparison group and a value of "1" to participants in the intervention group	Dichotomous	
Covariates — measured for intervention and comparison participants				
Age	How old are you?	10/11/12/13/14/15/16/17/18/19/20/21+	Ordinal	PREP pre-survey PREP post-survey

Table 3 (continued)

Variable	Question	Response categories	Operationalization	Source
Ethnicity	Are you Hispanic or Latino?	Yes/no	Dichotomous	PREP pre-survey PREP post-survey
Race	What is your race?	American Indian or Alaskan Native/Asian/Black or African American/Native Hawaiian or other Pacific Islander/White or Caucasian	Categorical	PREP pre-survey PREP post-survey
Gender	What is your gender?	Male/female/transgender male to female/transgender female to male/prefer not to answer	Dichotomized transgender Male to Female = female transgender Female to male = male Prefer not to answer = missing	PREP pre-survey PREP post-survey
Socioeconomic status (SES)	What is the highest grade or level of school completed by your mother or female guardian?	Less than high school graduate/GED/high school graduate/GED/more than high school graduate/GED	Categorical	PREP pre-survey
Knowledge of risky behaviors scale	Can the following behaviors put you at risk for HIV? a) Sharing needles for tattooing and piercing	Yes/no/not sure		PREP pre-survey
Knowledge of methods for prevention of pregnancy and STD scale	b) Having vaginal sex without a condom c) Donating blood* d) Using the same condom twice e) Hugging f) Having anal sex without a condom Which of the following methods are effective if used correctly to protect people from STDs (including HIV) and pregnancy? a) choosing not to have sex (abstinence) b) Using hormone based birth control (e.g., the pill, Depo-Provera shot, patch, vaginal ring) c) Using condoms d) Using withdrawal* e) Douching (washing out the vagina) * denotes items that were dropped in refined scale	Protects from pregnancy & STD/HIV/protects from pregnancy only/protects from neither	Continuous (mean scale score)	
Baseline sex	Have you ever had sexual intercourse?	Yes/no	Dichotomous	PREP pre-survey
Baseline pregnancy	To the best of your knowledge have you been pregnant or gotten someone else pregnant even if no child was born?	Yes/no	Dichotomous	PREP pre-survey
Sex past 3 months	In the past 3 months, with how many people did you have sexual intercourse, even if only one time?	0, I did not have sex/1 person/2–3 people/4+ people	Dichotomized to indicate "had sex in the past three months" 1: 1/2–3 people/4+ people 0: 0, I did not have sex	PREP pre-survey
Condom use past 3 months	When you had sexual intercourse in the past 3 months, how often did you or a partner use a condom?	All of the time/most of the time/some of the time/none of the time	Ordinal recoded to indicate protected sex in the past 3 months 1: All of the time 2: Most of the time/some of the time 3: None of the time 4: Did not have sex past 3 months	
Birth control use past 3 months	When you had sexual intercourse in the past 3 months, how often did you or a partner use birth control?	All of the time/most of the time/some of the time/none of the time	Ordinal recoded to indicate "protected sex in the past three months" 1: All of the time 2: Most of the time/some of the time	PREP pre-survey

Table 3 (continued)

Variable	Question	Response categories	Operationalization	Source
			3: None of the time	
			4: Did not have sex past 3 months	
Social-emotional competence	In the past 3 months, how often would you say you…	Five-point Likert	Ordinal	PREP pre-survey
	a) cared about doing well in school	All of the time/most of the time/some of the time/a little of the time/none of the time	a) recoded as	
	b) shared ideas or talked about things that really matter with a parent/guardian		1: All of the time	
			2: Most of the time	
	c) resisted or said no to peer pressure?		3: Some of the time	
	d) managed conflict without causing more conflict		4: A little/none of the time	

References

ATLASti. Version 1.0.24. (1999). Berlin, Germany: Scientific Sofware Development. Retrieved from: https://atlasti.com/.

Barrera, M., Berkel, C., & Castro, F. G. (2017). Directions for the advancement of culturally adapted preventive interventions: local adaptations, engagement, and sustainability. *Prevention Science, 18*, 640–648. https://doi.org/10.1007/s11121-016-0705-9.

Bell, S. G., Newcomer, S. F., Bachrach, C., Borawski, E., Jemmott III, J. B., Morrison, D., … Zimmerman, R. (2007). Challenges in replicating interventions. *Journal of Adolescent Health, 40*, 514–520. https://doi.org/10.1016/j.jadohealth.2006.09.005.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011a). Putting the pieces together: an integrated model of program implementation. *Prevention Science, 12*, 23–33. https://doi.org/10.1007/s11121-010-0186-1.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011b). Putting the pieces together: an integrated model of program implementation. *Prevention Science, 12*, 23–33. https://doi.org/10.1007/s11121-010-0186-1.

Berman, P., & McLaughlin, M. W. (1978). *Federal programs supporting educational change, Vol. VIII: Implementing and sustaining innovations.* Santa Monica: RAND Corporation, R-1589/8-HEW, 1978. Retrieved from: https://www.rand.org/pubs/reports/R1589z8.html.

Blake, S. M., Simkin, L., Ledsky, R., Perkins, C., & Calabrese, J. M. (2001). Effects of a parent-child communications intervention on young adolescents' risk for early onset of sexual intercourse. *Family Planning Perspectives, 33*, 52–61.

Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., & Emshoff, J. G. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15*, 253–268. https://doi.org/10.1007/BF00922697.

Bumbarger, B., & Perkins, D. (2008). After randomised trials: issues related to dissemination of evidence-based interventions. *Journal of Children's Services, 3*, 55–64.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007a). A conceptual framework for implementation fidelity. *Implementation Science, 2*, 40. https://doi.org/10.1186/1748-5908-2-40.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007b). A conceptual framework for implementation fidelity. *Implementation Science, 2*, 40. https://doi.org/10.1186/1748-5908-2-40.

Castro, F. G., Barrera, M., Jr., & Martinez, C. R., Jr. (2004). The cultural adaptation of prevention interventions: resolving tensions between fidelity and fit. *Prevention Science, 5*, 41–45. https://doi.org/10.1023/B:PREV.0000013980.12412.cd.

Century, J., Freeman, C., & Rudnick, M. (2008). Measuring and accumulating knowledge about fidelity of implementation (FOI) of science instructional materials. Retrieved from: http://cemsesite.cemseprojects.org/research-and-evaluation/research/foi/narst-framework.pdf .

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555.

Cross, W., & West, J. (2011). Examining implementer fidelity: conceptualising and measuring adherence and competence. *Journal of Children's Services, 6*, 18–33. https://doi.org/10.5042/jcs.2011.0123.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23–45.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350. https://doi.org/10.1007/s10464-008-9165-0.

Dusenbury, L. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237–256. https://doi.org/10.1093/her/18.2.237.

Ebbole, T. (2007). Evidence-based programs and practices: what does it all mean? Retrieved from: https://www.childhelp.org/wordpress/wp-content/uploads/2015/07/Research-Review-Evidence-Based-Programs-and-Practices-What-Does-It-All-Mean-Childrens-Services-Council.pdf.

Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science: The Official Journal of the Society for Prevention Research, 5*, 47–53.

Ennett, S. T., Haws, S., Ringwalt, C. L., Vincus, A. A., Hanley, S., Bowling, J. M., Rohrbach, L. A. (2011). Evidence-based practice in school substance use prevention: Fidelity of implementation under real-world conditions. *Health Education Research, 26*, 361–371.

Feldman Farb, A., Margolis, A. L. (2016). The teen pregnancy prevention program (2010-2015): Synthesis of impact findings. *American Journal of Public Health, 106*, S9–S15.

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—principles and practices. *Health Services Research, 48*(6 Pt 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117.

Firpo-Triplett, R., & Fuller T.. (2012). General adaptation guidance: a guide to adapting evidence-based sexual health curricula. ETR Associates and CDC Division of Reproductive Health. Retrieved from: http://recapp.etr.org/recapp/documents/programs/GeneralAdaptationGuidanceFINAL.pdf.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: a synthesis of the literature. Tampa, FL: University of South Florida, Louis de La Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231). Retrieved from: http://centerforchildwelfare2.fmhi.usf.edu/kb/Implementation/Implementation%20Research%20-%20A%20Synthesis%20of%20Literature%20%20-%202005.pdf.

Gottfredson, D. C., Gottfredson, G. D., & Hybl, L. G. (1993). Managing adolescent behavior a multiyear, multischool study. *American Educational Research Journal, 30*, 179–215. https://doi.org/10.3102/00028312030001179.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288.

Kam, C.-M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. Promoting alternative THinking skills curriculum. *Prevention Science: The Official Journal of the Society for Prevention Research, 4*, 55–63.

Kelsey, M., & Layzer, J. (2014a). Implementing three evidence-based program models: early lessons from the teen pregnancy prevention replication study. *Journal of Adolescent Health, 54*, S45–S52. https://doi.org/10.1016/j.jadohealth.2013.12.024.

Kelsey, M., & Layzer, J. (2014b). Implementing three evidence-based program models: early lessons from the teen pregnancy prevention replication study. *Journal of Adolescent Health, 54*, S45–S52. https://doi.org/10.1016/j.jadohealth.2013.12.024.

Kershner, S., Flynn, S., Prince, M., Potter, S. C., Craft, L., & Alton, F. (2014). Using data to improve fidelity when implementing evidence-based programs. *Journal of Adolescent Health, 54*, S29–S36. https://doi.org/10.1016/j.jadohealth.2013.11.027.

Margolis, A. L., & Roper, A. Y. (2014). Practical experience from the office of adolescent health's large scale implementation of an evidence-based teen pregnancy prevention program. *Journal of Adolescent Health, 54*, S10–S14. https://doi.org/10.1016/j.jadohealth.2013.11.026.

McGraw, S. A., Sellers, D. E., Johnson, C. C., Stone, E. J., Bachman, K. J., Bebchuk, J., … Edmundson, E. W. (1996). Using process data to explain outcomes an illustration from the child and adolescent trial for cardiovascular health (CATCH). *Evaluation Review, 20*, 291–312. https://doi.org/10.1177/0193841X9602000304.

McHugh, R. K., Murray, H. W., & Barlow, D. H. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: the promise of transdiagnostic interventions. *Behaviour Research and Therapy, 47*, 946–953. https://doi.org/10.1016/j.brat.2009.07.005.

Metz, A., Naoom, S. F., Halle, T., & Bartley, L. (2015). *An integrated stage-based framework for implementation of early childhood programs and systems (OPRE Research Brief OPRE 201548)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Mmari, K. N., Blum, R. W., Organization, W. H., et al. (2005). Risk and protective factors affecting adolescent reproductive health in developing countries. Retrieved from: http://apps.who.int/iris/handle/10665/43341.

Moore, J. E., Bumbarger, B. K., & Cooper, B. R. (2013). Examining adaptations of evidence-based programs in natural contexts. *The Journal of Primary Prevention, 34*, 147–161. https://doi.org/10.1007/s10935-013-0303-6.

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research, 39*, 374–396. https://doi.org/10.1007/s11414-012-9295-x.

Philliber, S. (2015). Evaluating teen pregnancy prevention programs: decades of evolving strategies and practices. *Societies, 5*, 631–645. https://doi.org/10.3390/soc5030631.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41–55. https://doi.org/10.1093/biomet/70.1.41.

Schoenfelder, E. N., Sandler, I. N., Millsap, R. E., Wolchik, S. A., Berkel, C., & Ayers, T. S. (2013). Caregiver responsiveness to the family bereavement program: what predicts responsiveness? What does responsiveness predict? *Prevention Science: The Official Journal of the Society for Prevention Research, 14*, 545–556. https://doi.org/10.1007/s11121-012-0337-7.

Spoth, R., Guyll, M., Trudeau, L., & Goldberg-Lillehoj, C. (2002). Two studies of proximal outcomes and implementation quality of universal preventive interventions in a community-university collaboration context. *Journal of Community Psychology, 30*, 499–518. https://doi.org/10.1002/jcop.10021.

State personal responsibility education program fact sheet | Family and Youth Services Bureau | Administration for Children and Families. (n.d.). Retrieved September 16, 2015, from: http://www.acf.hhs.gov/programs/fysb/resource/prep-fact-sheet.

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., et al. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ, 338*, b2393–b2393. https://doi.org/10.1136/bmj.b2393.

Stirman, S., Miller, C. J., Toder, K., & Calloway, A. (2013). Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implementation Science, 8*, 65. https://doi.org/10.1186/1748-5908-8-65.

Stuart, E. A., & Green, K. M. (2008). Using full matching to estimate causal effects in nonexperimental studies: examining the relationship between adolescent marijuana use and adult outcomes. *Developmental Psychology, 44*, 395–406. https://doi.org/10.1037/0012-1649.44.2.395.

Wamoyi, J., & Wight, D. (2014). 'Mum never loved me.' How structural factors influence adolescent sexual and reproductive health through parent–child connectedness: a qualitative study in rural Tanzania. *African Journal of AIDS Research, 13*, 169–178. https://doi.org/10.2989/16085906.2014.945387.

Wang, B., Deveaux, L., Knowles, V., Koci, V., Rolle, G., Lunn, S., Li, X., Stanton, B. (2015). Fidelity of Implementation of an Evidence-Based HIV Prevention Program among Bahamian Sixth Grade Students. *Prevention Science, 16*, 110–121.