Appendix: Supplemental Online Content

A Randomized Controlled Trial of Dating Matters®: Effects on Teen Dating Violence and Relationship Behaviors

Phyllis Holditch Niolon, PhD1, Alana M. Vivolo-Kantor, PhD, MPH1, Allison J. Tracy, PhD2, Natasha E. Latzman, PhD1, Todd D. Little, PhD3, Sarah DeGue, PhD1, Kyle M. Lang, PhD3, Lianne Fuino Estefan, PhD, MPH1, Sharon R. Ghazarian, PhD2, Wendy LiKamWa McIntosh, MPH1, Bruce Taylor, PhD4, Linda Johnson5, Henrietta Kuoh, MPH1, Tessa Burton, MPH1, Beverly Fortson, PhD1, Elizabeth A. Mumford, PhD4, Shannon Nelson, MA4, Hannah Joseph4, Linda Anne Valle, PhD1, Andra Teten Tharp, PhD1

Content:

Appendix Supplemental Text

Appendix Table 1. Full Sample Demographics and Scores on Teen Dating Violence Perpetration and Victimization, Negative Conflict Resolution Strategies, and Positive Relationship Behaviors

Appendix Table 2. Test of Baseline Equivalence on Race, Ethnicity, and Age

Appendix Table 3. Intervention Components by Condition

Appendix Table 4. Measure of Teen Dating Violence

Appendix Table 5. Measure of Negative Conflict Resolution Strategies

Appendix Table 6. Measure of Positive Relationship Skills

 Appendix Table 7a. Means of Outcome Variables Using Item Averages (Original Metric)

Appendix Table 7b. Means of Outcome Latent Variables (POMS Metric)

Appendix Table 8. Cronbach’s Alpha Reliability Coefficients

Appendix Table 9. Conversion of Original Response Anchors to POMS Scores

Appendix Table 10. Unstandardized Measurement Model Results

Appendix Table 11. Results of Model Tests

Appendix Table 12. Program Effect Size Estimates (Cohen’s *d* statistic)

Appendix Figure 1. Data Assessment Design

Appendix Figure 2. Relative Risk Reduction for DM vs. SC by cohort, gender, and time periods: Teen Dating Violence Perpetration

Appendix Figure 3. Relative Risk Reduction Relative Risk Reduction for DM vs. SC by cohort, gender, and time periods: Teen Dating Violence Victimization

Appendix Figure 4. Relative Risk Reduction Relative Risk Reduction for DM vs. SC by cohort, gender, and time periods: Negative Conflict Resolution

Appendix Figure 5. Percent relative risk reduction (M, range) for DM vs. SC across cohorts and time periods

Appendix Figure 6. Constrained Means Across Time by Sex and Cohort – Positive Relationship Skills

**Appendix**

***Design and Sample:***

The *Dating Matters: Strategies to Promote Healthy Relationships (Dating Matters;* DM*)* Initiative employed a multi-site, cluster randomized controlled trial to evaluate the comparative effectiveness of the *Dating Matters* comprehensive prevention model (DM) relative to an existing, evidence-based teen dating violence prevention curriculum, which served as the standard-of-care (SC) comparison condition. The research trial was conducted by CDC through a contract with NORC at the University of Chicago. NORC assisted CDC with design and measurement development and conducted the data collection. Implementation of the two models of teen dating violence prevention, DM and SC, were conducted by funded local health departments who partnered with schools and communities in high-risk urban neighborhoods.

***Site selection, school recruitment, random assignment, and attrition.***In 2010, CDC released a funding opportunity announcement for local health departments (LHDs) to implement a teen dating violence (TDV) prevention initiative entitled *Dating Matters: Strategies to Promote Healthy Teen Relationships* (CDC-RFA-CE11-1103). Local health departments in cities with metropolitan statistical areas of over 1 million people were eligible to apply. The funding opportunity announcement required that LHDs identify and partner with 10-12 neighborhood middle schools (6th to 8th grades) in “high-risk” communities (defined as having higher than average crime and economic disadvantage) within their city or county to participate in the *Dating Matters* Initiative for five years (1 planning year and 4 years of implementation). High-risk, urban communities were chosen as the targeted population for this study because research indicates that youth living in high-risk, urban settings are likely to be exposed to multiple risk factors for TDV, such as community violence and economic disadvantage,1-6 potentially putting them at higher risk for TDV. However, few TDV prevention programs have been tested in this population. The announcement indicated that sites would be part of a multi-site evaluation and that CDC and its partners would randomly assign schools from their site to one of two TDV prevention strategies; funded sites agreed to cooperate with the evaluation.

NORC used a computer-generated random numbers approach to assign an equal number of schools in each site to either the DM or SC condition. NORC conducted randomization independently in each site/district using a simple random assignment process (i.e., the parallel to flipping a coin) such that each school within each site/district had an equal chance of being selected for either the DM or SC condition. Simple random assignment had several advantages. First, it was easy to implement and was easily understood by stakeholders. These features were important to assure buy-in locally as to the fairness and scientific rigor of the evaluation design. Also, simple random sampling yields data that can be easily analyzed without concern for additional weighting and loss of statistical power. A second approach of randomizing schools within “blocks” (i.e., groups) of similar schools was considered. However, in the analysis of publicly available data for each participating school on characteristics that have been found to be important when delivering interventions in the past (e.g., school size, student-teacher ratios, percent receiving free lunch, etc.) no substantial differences among the schools were found to warrant blocking overall. The random assignment process used in each site was documented using screen shots of each step of the randomization program which were placed into a site specific memo and were made available for auditing of the integrity of the process. The randomization program, based on a script developed for use in Microsoft Excel, assigned a random number between zero and one for each school. From all of the assigned random numbers a median score (the 50th percentile) was calculated. Any schools that received a random number below the median score were assigned to SC and and those above the median score were assigned to DM. CDC and NORC held calls with the LHDs to inform them of the random assignment of their schools and review their site specific randomization memo on the procedures used. These calls ensured that they understood the random assignment process and would agree to comply with its results. The LHDs then informed the schools of their assignment. As described below and shown in the CONSORT figure, some schools dropped out of the study and were replaced by alternate schools. Each alternate school needed to meet the same criteria as the original schools (e.g., economically disadvantaged) and was randomized at the site level using a similar randomization process as described above.

Forty-six schools were initially identified for participation by the 4 sites (12 schools in 3 sites and 10 schools in one site). Because some schools closed or dropped out of the study, replacement schools were recruited during the first three years of the study. Not counting school closures (n = 4 schools that closed before minimum level of participation), overall school-level attrition was 20.69% (12 of 58 schools); 20.00% of DM schools and 21.43% of SC schools failed to participate after recruitment and randomization or left the study after some participation but were active less than two years (differential attrition = 1.43%). Schools who implemented either DM (n=22) or SC (n=24) for at least two full academic years (see Figure 1 in the manuscript) were included in the analyses. The decision to include schools for analysis based on two full years of participation in the trial was based on the fact that schools implementing less than two years would have implemented less than half of the 3-year middle school span covered by the DM components and that students from the schools would have less than half of the survey data collection opportunities across the three years of middle school. Based on Cox and Hedge’s *g* cutoffs (0.05 and lower indicates equivalence),7 included schools were similar to excluded schools with respect to the racial/ethnic composition of the student body and the proportion of free or reduced-price lunch. Schools excluded from the study had a smaller average student body size (M = 236.97, SD = 180.72) than those that were included (M = 397.73, SD = 387.42). Further, excluded schools had a lower student-teacher ratio (M = 15.24, SD = 1.52) than included schools (M = 17.86, SD = 2.84).

***Sample.*** Data were collected from 5 cohorts of middle school students (N=7,847) from 2012-2015. Cohorts 1-3 were in 8th, 7th, and 6th grades, respectively, in the 2012-2013 school year. Cohorts 4 and 5 were added as 6th graders in 2013-2014 and 2014-2015, respectively (See Appendix Figure 1). The study is continuing to follow this sample of students as they matriculate into high school, through the end of the 2017-2018 school year; high school data will be analyzed once collection is complete. Data collected from parents and educators is not reported in this analysis. This paper presents outcome data for the two full-exposure cohorts of middle school students (Cohorts 3 and 4) who, in the DM condition, had the potential for exposure to all three grades/years of the DM model during the implementation period (2012-2016). Local ethics review boards at each of the four sites required active parental consent for research participation; parents were asked to sign and return forms indicating whether or not they gave permission for their child to complete surveys. Due to significant difficulty obtaining the minimum form return rate of 60% in one site, their review board allowed a switch to a passive parental consent procedure in that site starting in the second year of implementation (the 2013-2104 school year). The overall form return rate across sites was 74%, with 78% of those returning forms indicating that they gave permission for their child to participate in the trial. In three sites, schools typically had 3-4 classrooms of students per grade and all students in each grade were recruited into the study sample. One site had large schools, with an average of 10-11 classrooms per grade. Due to budget constraints, all students in each grade could not be recruited in these schools. Therefore, four classrooms per grade in these schools were randomly selected for recruitment into the trial; however, all students in these schools were still exposed to their school’s designated curricula (either DM or SC).The analytic sample for the current paper includes students from Cohorts 3 and 4 from schools who had implemented either DM (n=22) or SC (n=24) for at least two full academic years and who reported having dated at some point during middle school (N=2,349 students; DM: n = 1,157; SC: n = 1,192). Sample demographics and average scores on outcome variables are presented in Appendix Table 1. Demographic equivalence by condition was assessed within each cohort by gender (See Appendix Table 2) using a dataset that contains the grand mean imputed values drawn from 100 imputed datasets (see Measurement section for details). Differences by race were seen for some cohorts. There were more DM black boys in Cohort 3 (DM = 60.3%, SC = 52.6%) and fewer DM Hispanic boys in both cohorts (DM = 26.8%, SC = 35.2% for Cohort 3; DM = 24.2%, SC = 29.4% for Cohort 4). Other baseline differences in student racial/ethnic categories were small (< 5%), posing low risk for sample bias.

***Dating Matters condition (DM).*** The *Dating Matters* comprehensive prevention model adds to existing knowledge on TDV prevention in that it: 1) addresses the prevention of TDV and multiple risk factors for TDV with a comprehensive set of interventions targeting multiple levels of the social ecology; and 2) was developed for middle school students to prevent TDV before it starts by intervening at the developmental stage when dating typically initiates. The *Dating Matters* model8,9 includes the following interventions:

Youth programs. The 6th and 7th grade youth programs were developed by CDC. Eighth grade students received *Safe Dates*,10 an evidence-based dating violence prevention program. All three curricula use social-emotional learning and skills-based approaches to focus on healthy relationships and help youth learn and practice healthy relationship skills, such as communication and conflict resolution. The 6th grade curriculum includes 6 classroom sessions focusing on healthy relationships broadly, while the 7th grade (7 classroom sessions) and 8th grade curriculum (10 classroom sessions, a poster contest, and a play) are more focused on dating relationships. DM also includes an integrated, youth-delivered and youth-focused communications program, *i2i: What R U Looking 4,* designed to reinforce messaging from the DM curricula for all grades through a near-peer brand ambassador program, in-person activities, youth-developed materials, and digital resources led by DM brand ambassadors, who are volunteer high school students from the same neighborhoods as students in the DM middle schools.

Parent programs. In the DM condition, parents of 6th graders received a CDC-adapted version of the *Parents Matter!* program, an evidence-based program for decreasing sexual risk and promoting positive parent-child communication about sexual health among adolescents.11 Parents of 7th graders received a parent training program developed by CDC specifically for DM that was focused on positive parenting and parent-child communication. Parents of 8th graders received *Families for Safe Dates*, an evidence-based dating violence prevention intervention.12 Parent training was provided in community-based settings and not offered to parents of youth in the SC condition.

School and community-level interventions. All educators in DM schools were asked to complete an online training developed by CDC to provide educators and school personnel with TDV knowledge and resources and to motivate them to implement prevention activities in their schools. Local health departments implementing DM also engaged in a capacity and readiness assessment using a CDC-developed tool called the *Dating Matters Capacity and Readiness Tool* (DM-CAPT), as well as activities at the community-level aimed at tracking and informing local policy and data use related TDV prevention; these activities may have impacted students in both the DM and SC schools. See Appendix Table 3 for further depiction of the DM intervention compared to the SC condition. See [www.cdc.gov/violenceprevention/datingmatters](http://www.cdc.gov/violenceprevention/datingmatters) for more information on the DM component interventions.

***Standard of Care condition (SC):*** Students in the standard of care condition received *Safe Dates*,10 an evidence-based dating violence prevention program, in the 8th grade. No other interventions were implemented in the SC condition.

***Development of the Dating Matters Comprehensive Prevention Model (for more information see Tharp et al 2011)9:*** Findings from the broader violence prevention field (e.g., youth violence) suggest that multi-component, comprehensive prevention approaches that target multiple levels of the social ecology (individual, relationship, community, and society) are more effective than single-component approaches targeting only one level.11,12 When DM was being developed in 2009, two TDV prevention programs had evidence for effectiveness on TDV perpetration outcomes: *Safe Dates®*10 and *the Fourth R.*13 Both were school-based programs targeting mid-adolescents with individual level (i.e., targeting the students directly), skills-based interventions*.* The goal of DM was to build upon the existing evidence to develop a comprehensive primary prevention approach with the potential for greater and sustained impact on TDV perpetration, TDV victimization, and healthy relationship behaviors.

Toward this end, a whole middle-school approach was employed with youth programs in 6th, 7th, and 8th grades. *Safe Dates* was chosen as the evidence-based program for 8th grade students, and it informed development of complimentary content of the CDC-developed 6th and 7th grade programs. *Safe Dates* was selected because it had evidence of (1) primary and secondary prevention effects (e.g., preventing continued TDV among those already experiencing it, and preventing those who had no experience with TDV from experiencing it); (2) effects on perpetration and victimization of TDV; (3) effectiveness for boys and girls; and (4) sustained effects at 4-year follow-up.10 DM 6th and 7th grade youth program development was led by a CDC scientist with expertise in adolescent development. The programs are based on the best available empirical and experiential evidence, heavily influenced by feedback and expertise provided at each stage of the development process (i.e., expert review by scientists and practitioners; pilots at four independent sites; revision based on feedback). Both 6th and 7th grade programs focus on emotion-regulation and social skills within relationships. The 6th grade program focuses on relationships more broadly (e.g., peers, parents, dating partners), and while the 7th grade program addresses dating relationships more directly and includes additional content on sexual consent and coercion.

Parents play an important role in the lives of their adolescents as both teacher and role model. Certain parenting behaviors (e.g., inadequate supervision; lack of parental warmth; inconsistent or harsh discipline; exposure to family violence) are associated with TDV. Thus, a program that aims to improve parenting practices and the parent-child relationship in early adolescence help provide parents with the skills they need to help their children form healthy relationships.14-19 As such, parents are the focus of a second, relationship-level programmatic component of DM*:* parent programs for parents of 6th, 7th, and 8th graders*.* Evidence-based programs were chosen when available. Parents of 6th graders were offered an adapted version of *Parents Matter!*, an evidence-based prevention program designed to enhance protective parenting practices and encourage parent/child communication about sexuality and sexual risk behaviors.20,21 In consultation with the program developer, an additional session on healthy dating relationships and TDV was added to the program for use in DM. Parents of 8th graders were offered *Families for Safe Dates®*16*,* which is a booklet-based in-home teen dating violence prevention program completed by parents and their children. This evidence-based program provides families six booklets for parents and their teens to explore together to learn about different topics regarding TDV. The 7th grade parent program was evidence-informed and focused on parent-adolescent communication and positive parenting to reduce conflict, solve problems, and enhance the parent-child relationship. It was a combination of in person, group sessions, and parent-led, at-home sessions developed to complement and reinforce content from the 6th and 8th grade programs.

Schools represent one of the most salient and influential community contexts for adolescents. The *Dating Matters* Training for Educators was developed through a CDC partnership with Liz Claiborne, Inc. as an online training for teachers and other individuals who work with youth, such as school personnel, coaches, and youth mentors, and serves as the school-level component of the DM model. The training covers understanding TDV and its consequences, risk factors for TDV, and the importance of healthy relationships across the lifespan. The program is designed to provide educators with knowledge about TDV and resources related to prevention, and motivates them to implement prevention measures in their schools.9

CDC developed *i2i: What R U Looking 4?*, a youth-led, near-peer communications program delivered outside of school by older teens within students’ neighborhoods ,as the neighborhood/community component of the DMmodel. CDC developed *i2i* based on relevant behavior theory (e.g., Social Learning Theory, Diffusion of Innovation, Theory of Planned Behavior), and formative research. Formative research was conducted to: 1) incorporate youth voice by exploring knowledge, attitudes, and behaviors of youth ages 11 to 14 toward dating and relationships; 2) test potential communications messages and concepts with youth; 3) survey demographically similar communities on capacity to implement a youth communications program; 4) analyze a youth-focused audience; and 5) review the literature on communication campaigns geared towards similar audiences on sensitive topics and key influencers in the lives of high-risk, urban youth. Additionally, CDC consulted with experts, which included audience and industry experts from diverse backgrounds.

As additional community-level components, local health departments were provided with CDC-developed tools for assessing and tracking capacity and readiness to do TDV prevention work, and for assessing and tracking local policies that might be relevant to TDV prevention. CDC scientists with expertise in program implementation and technical assistance developed the capacity and readiness tool in collaboration with health department staff. CDC scientists with expertise in policy analysis and assessment developed a policy toolkit, a collection of tools to help health departments and their partners learn about policy assessment and analysis.9 These components were intended to help facilitate community-level changes in capacity for TDV prevention and local policy that could support and reinforce efforts with youth, parents, and schools.

***Implementation of the conditions:*** LHDs were responsible for implementation of the two conditions, and their non-research cooperative agreement funding mechanisms was not prescriptive about how they funded or executed implementation. Therefore, there was some variation in who implemented the programs, although the conditions were implemented with the same content and structure across sites. In Alameda, a combination of health department staff and trained volunteers from a national civil society program implemented student curricula; health department staff implemented the parent curricula, and a youth-led, nonprofit media production company assisted with the implementation of the communications program. In Baltimore, health department staff and trained volunteers implemented both the student and parent curricula, and research staff from a local universityassisted with the implementation of the communications program. In Broward, teachers implemented the student curricula; health department staff implemented the parent curricula, and a communications specialist from the health department assisted with the implementation of the communications program. In Chicago, staff at a contracted social service agency implemented student curricula and the parent curricula and assisted with the communications program; health department staff also assisted with the communications program. Although the sites varied in their implementation, implementation of the student curricula did not differ across conditions within a site. For instance in Broward, teachers implemented curricula in both conditions, and in Chicago, the same contractors implemented student curricula in both conditions.

***Student survey data collection.*** Paper-and–pencil scannable surveys were administered twice during the academic year (fall and spring) in a classroom setting. The assessment of cohorts by grade and school year is illustrated in Appendix Figure 1. All students who had positive parental consent were invited to participate in a survey; all students who assented were surveyed. Surveys were kept confidential and were marked only by a tracking ID number. Make-up survey days were conducted in most schools to try to give students who were absent on the original survey day an opportunity to participate. Copies of the student surveys are available at <http://www.norc.org/Research/Projects/Pages/CDC-Dating-Matters-Experimental-Evaluation.aspx> under the resources tab.

***Implementation evaluation data collection.*** Though data is not presented in this report, a process evaluation was also conducted for both the student programming and parent programming. From August 2012 to August 2016, five types of process evaluation data were collected: (1) student session attendance for students with positive parental consent; (2) session logs for the 6th, 7th, and 8th grade student curricula, which enabled implementers to document fidelity and changes made to each session, implementation challenges, and student engagement and understanding; (3) session logs for the 6th and 7th grade parent curricula, which enabled facilitators to document fidelity and changes made to each session, facilitation challenges, and parent attendance, engagement, and understanding; (4) counts reported by sites for 8th grade *Families for Safe Dates* completion, which enabled reporting on parent participation in this program; and (5) parent satisfaction surveys, which were used to collect feedback from parent participants about their experiences in the program.

In September 2013, sites began using paper-and-pencil scannable forms for all session logs and parent satisfaction surveys. Each student implementer and parent facilitator completed hard-copy session logs for the student and parent curricula using paper-and-pencil scannable forms provided by the implementation evaluation contractor (RTI). RTI provided shipping materials to each site, and sites were asked to submit the session logs to RTI by FedEx on a weekly basis. Parent satisfaction surveys were completed by participating parents at the end of each implementation cycle using scannable paper-and-pencil questionnaires, which were also shipped back to RTI. Analyses of implementation data is currently underway, and results are not yet available.

A total of 13,791 student session logs (9,090 were DM and 4,701 were SD) were collected during the initiative (August 2012 to August 2016) across 1,678 classrooms (1,204 were DM and 474 were SD). Using the highest reported number of participants during the cycle for each class, it is estimated that 36,890 students participated in teen dating violence prevention programs during the study – 25,808 students participated in the DM programs and 11,082 participated in the SC program. Attendance at parent programs was also collected. Across the study (August 2012 to August 2016), a total of 791 parents participated in the DM in-person parent programs (6th and 7th grade parents); 176 parents completed the *Families for Safe Dates* at-home program (8th grade parents), for a total of 967 parent participants overall.

CDC also collected cost data on the implementation of DM. A cost analysis is forthcoming.

**Missing Data**

Although the analysis sample for this paper includes only a subset of sample participants, the entire dataset was used for missing data treatment. Because of the way the data were aligned for analyses and other unplanned reasons (e.g., attrition), there was a high proportion of missing data throughout, typically 65% or more. Some data were missing by design (e.g., Cohort 1 was not measured in sixth or seventh grade, Cohort 2 was not measured in sixth grade, and Cohort 5 was not measured in eighth grade). Data missing by design were considered missing completely at random (MCAR), because these missing data arise from a random process controlled by the researcher. School-level missing data due to attrition (early exit) and the study’s attrition-responsive recruitment strategy (late entry) and missing data due to student-level processes (nonresponse due to school transfer, absenteeism, refusal, spot missingness within a survey, or unclear/unusable responses) were treated as missing at random (MAR). The MAR assumption requires that the causes of nonresponse are captured by the observed portion of the dataset. This assumption is quite reasonable for a large-scale data collection such as *Dating Matters*, given the breadth of measures that were included in the overall protocol.22 No definitive test for the MAR assumption exists,23 but the missing data treatment employed in this study was designed to capture any extant MAR process.24

When key variables are associated with a high nonresponse rate, modern methods for handling missing data are required to preserve the available information. In this study, multiple imputation was used, which is considered state-of-the-art among methods designed to avoid the bias that case deletion typically introduces. To increase the precision of the imputed data estimates, the imputation models drew from survey responses obtained from both the study sample and students excluded from analysis (Cohorts 1, 2, and 5, and students from schools implementing less than 2 years), as well as school-level information.

All missing data were multiply imputed using PcAux,25 which was programmed in R.26 The PcAux package implements the ideas of Howard, Rhemtulla, and Little24 to create principled multiple imputations (MIs) via sequentially employed principal components regression as the elementary imputation method. Additional details on the algorithms implemented by PcAux are available in Enders;23 Little, Jorgensen, Lang, and Moore;27 and Lang and Little22 for reviews of MI and its relative strengths. Further information on the sequential regression approach for MI is available in Van Buuren28 and Van Buuren, Brand, Groothuis-Oudshoorn and Rubin.29 Hastie, Tibshirani, and Friedman30 provide an introduction to principal components regression.

In broad strokes, MI operates by replacing missing data with predicted values from a stochastic regression equation in which the incomplete variables being imputed act as the dependent variable and other variables on the dataset act as predictors. This process is repeated many times (e.g., *m*=100) so that each missing value is represented by a distribution of plausible replacements. The variability of these distributions (i.e., the *between-imputation variance*) provides an indication of how much information has been lost to nonresponse and controls for uncertainty in the true values of the missing data by increasing sampling variability in parameters estimated from the multiply imputed data.31

MI can be computationally intensive and difficult to implement with large datasets. The PcAux package is designed to mitigate this computational expense without compromising the imputations’ quality. This goal is achieved by summarizing the incomplete data with a set of principal component scores and using those scores as predictors in the regression equations used to create the imputed values—in other words, principal components regressions (PCRs) are used as the imputation models. Imputing with PCRs—as opposed to the traditional approach that fits the imputation models to the raw variables—allows analysts of missing data to use a larger portion of the information in the data to create the imputations. In the traditional approach to large datasets like the *Dating Matters* data, only a small fraction of the potential predictors can be included in the imputation models, and choosing which predictors to include can be a difficult and highly subjective task. The PcAux approach largely bypasses this issue by reducing the entire dataset into principal component scores and providing empirical criteria to inform the number of scores to use as predictors (the proportion of variance in the raw data explained). The imputation models employed 421 component scores, which explained 50% of the variance in the data.

The computational efficiencies achieved by employing PCRs as the imputation models also allow analysts of missing data to incorporate more nonlinear terms into the imputation models’ predictor set. When the inferential analysis will include nonlinear effects (e.g., interactions), these nonlinear effects must also be included as predictors in the imputation models.32 In the traditional approach, including interaction terms only increases the size of an already too large pool of potential predictors. In contrast, the PcAux approach can accommodate many interaction terms, even with large datasets, by collapsing them into a secondary set of principal component scores. For the current analysis, the two-way interactions between each variable and student sex, student ethnicity, treatment condition, cohort, site, and school were included in the imputation process. The models employed 326 interaction component scores, in addition to the 421 main component scores. These additional components explained 10% of the variance in the set of interactions.

**eMeasures**

Item wording and response scales for all outcome measures are presented in Appendix Tables 4-6. Scale means obtained by averaging the raw items from the imputed grand mean dataset are given in Appendix Table 7 as an indication of the prevalence of these behaviors. Cronbach’s alpha reliability coefficients are given in Appendix Table 8, although item average composite variables were not used in analysis (see Item Parceling section below).

***Item Parceling.*** When possible, a latent variable approach was used to maximize the precision of the estimates of program impact. Item parcels (mean values of item subsets) were used as indicators of the latent constructs because parcels have higher reliability, greater communality, higher ratio of common-to-unique factor invariance, lower likelihood of distributional violations, and a greater number of intervals that are tighter and more equally spaced. Parceling requires fewer parameter estimates and has a lower indicator-to-subject ratio, lower likelihood of correlated residuals and dual factor loadings, and reduced sources of sampling error.33-35 Selection of items in this study reflects “facet-representative parceling.” In other words, indicators are created from correlated items representing coherent dimensions of a construct rather than items drawn from multiple dimensions. This strategy ensures that content validity is retained and that correlated errors do not inflate the factor loadings and reduce the ability to parse error from true-score variance. All outcomes in this study were represented by latent variables with three parceled indicators.

Common conceptualizations of the TDV construct depict an overarching general construct that may be used as a total score, as well as the specific subscales that comprise the larger construct.36,37 Existing research typically employs either the total score or the subscale method to examine TDV as a variable of interest.38-40 Conclusions vary, however, across studies that report findings based on the general TDV construct or specific subscales.

Because this construct contains such a large set of items, preliminary measurement development was conducted using data from the first 2 years of the study to examine the structure of the items and to determine how they might be best modeled in the outcome assessment. A summary of the measurement development work is reported here. Detailed information can be obtained upon request. The item set was examined using a bifactor modeling approach,41 which is particularly useful because it differentiates the common and specific variance of the items themselves. All measured items are loaded onto both the general and specific constructs. In general, higher loadings on a common factor than on domain-specific factors suggest that items may be more salient to the common construct. Conversely, higher loadings leading to the domain-specific construct suggest salience for a specific domain instead of the common construct (see Gomez and McLaren42 for an example).

Separate bifactor models of the teen dating violence items were examined for perpetration and victimization data. Models were compared across gender, ethnicity, and grade. Results demonstrated that subscales contributed significant variance to the model, but that common variance for the overarching TDV construct was more salient. These results replicated across both perpetration and victimization item sets. Models demonstrated good fit across four waves of data.

Based on these findings, it was concluded that the TDV subscales do not give strong enough information, above and beyond the common construct, to include separately in the outcome assessment models. However, items in specific subscales contained a substantial amount of variance shared within the subset of items, and different subscales correlated more highly with one another than other subscales, suggesting that the subscale scores can be used as indicators of the general TDV construct using latent variable modeling.

#### ***POMS Scoring.*** Since outcomes were measured with a different number of items and a different number and wording of response categories, the parceled outcome indicators were standardized to reflect a “percent of maximum score” (POMS), rescaling the original metric to range from 0 to 100 using the equation ((X – 1)/4)\*100; see Little33). This is a convenient way of anchoring the scales in a way that makes descriptive comparisons across outcomes possible, although no statistical tests of differences in effects were conducted across outcomes. Appendix Table 9 shows how the original scales were recoded to reflect POMS scaling.

**Data Preparation**

***Covariate Adjustment.*** To reduce the complexity and estimation load of the impact assessment models, a method designed to statistically adjust parcels for covariate effects prior to latent variable modeling (see Little33 and Lance43 for a discussion of this approach) was used. Specifically, the parcel indicators of each outcome were regressed separately on a set of covariates and calculated new variables capturing individual-level deviations of observed parcel scores from model-based predicted values. This pre-analysis data preparation approach is most effective when many covariates are involved; the statistical adjustments from many covariates will reduce measurement error without concern for issues related to multicollinearity or statistical evidence of independent effects. These covariate-adjusted variables were used as the latent variable indicators in the impact assessment models. Many indicators used in this study could best be described by mixture distributions with a heavy positive skew and a preponderance of zeros. For these parcels, a generalized linear regression model for a Tweedie distribution with power parameter of 1.5 and a log link was used. Other parcels were more normally distributed and so a linear regression was used for covariate adjustment. The models were estimated with data stacked by time, so that a single regression coefficient was estimated for each covariate regardless of time point. This ensured measurement invariance in the covariate-adjustment process across groups and time.

Covariates were coded so that the intercept of the covariate-adjustment model reflects the sample. For example, continuous covariates (e.g., relative age) were grand-mean centered. Traditional dichotomous coding was used for nominal variables with an interpretationally desirable anchor. For example, a 4-month recall window was chosen as the zero-coded category to statistically adjust downward the inflated values obtained from a lifetime recall window. For categorical variables without a meaningful anchor category (e.g., school membership), a non-standard[[1]](#footnote-1) effects-coded parmeterization44 was used. In this approach, each category contributes proportionally to the covariate adjustment, yielding an intercept that reflects the covariate-scrubbed level of the full sample.

Effects-coded covariates in a regression model scales the resulting intercept in a way that reflects the makeup of the full sample cast into levels scrubbed of covariate influence. Effects-coded scaling of observed variables is accomplished by creating a set of indicators for each categorical variable, one fewer than the number of response categories. This method parallels traditional dummy-coding but assigns three values: -1, 0, or 1, rather than two: 0 or 1. Where dummy-coding returns a 0 value on all indicators of a categorical variable for students in the reference category, no student receives a full set of 0s for effects-coded indicators of a categorical variable. Instead, a student belonging to an arbitrarily chosen category receives -1 for all indicators. All other students receive 0s for all but one indicator; for the indicator representing the category to which they belonged, student received a code of 1.

While the parcels were calculated to reflect a POM scale, the covariate-adjusted residual scores do not have meaningful 0 and 100 values that can anchor the POM boundaries. The covariate-adjustment model implies that there is a distribution of residuals around each calculated parcel score. For example, when the observed score for an outcome is zero, the distribution of residuals around this value results in some negative values. In other words, residualizing produces an observed distribution of covariate-adjusted scores that can fall outside the bounds of 0 and 100. To return to a meaningful approximation of POM, the covariate-adjusted values were parameterized to identify two points along the underlying continuum that could be used to anchor the 0 and 100 boundaries, following the logic below.

The predicted value of the outcome score calculated for a hypothetical “average” student (i.e., one for whom all covariates are zero) will always take the same value, regardless of the student’s observed score on the dependent variable. However, the covariate-adjusted or residualized score is different for students identical with respect to covariate characteristics but with different observed scores. For example, a prototypical student (all covariate values are zero) who reported the lowest value of the observed parcel score would receive a residualized value equal to 0 minus the intercept. A prototypical student who reported the highest value of the observed parcel score would receive a residualized value equal to 100 minus the intercept. This provides a meaningful anchor at either end of the POM scale against which to make inferences about effect sizes, despite calculated values that lie outside the admissible boundary values of 0 and 100.

The nested structure of the sample was handled in this study by including a set of zero/one indicators of school membership as covariates. Because these indicators capture the stratification of the sample by site and all between-school characteristics, including school size, student-teacher ratio, and racial/ethnic composition, no other school-level covariates were included in the covariate adjustment model. While this approach to nested data is an indirect correction, it is particularly justified when a multiple group latent variable approach is preferred to address treatment effects. Further, the means structure of the data is efficiently adjusted for any school-specific variation.

All students received the Baseline version of the survey upon their school’s entry into the study. The timeframe reference for some items differed across the Baseline and Follow-Up versions of the survey; the Baseline version referred to lifetime behaviors and the Follow-Up version referred to behaviors in the past 4 months. The survey version was coded 0=Follow-Up and 1=Baseline so that the covariate-adjusted outcomes reflect behaviors that were standardized to a 4-month recall period.

Witnessing violence was measured by the Juvenile Victimization Questionnaire – Child Self-Report Version.45 For this study, two time-varying dichotomous indicators were used: violence involving a parent in an intimate partner relationship (did you see or hear a parent get pushed, slapped, hit, punched, or beat up by another parent, or their boyfriend/girlfriend) and violence in the community (did you see anyone get attacked on purpose with a stick, rock, gun, knife, or other thing that would hurt at home, at school, at a store, in a car, on the street, or anywhere else; were you in any place in real life where you could see or hear people being shot, bombs going off, or street riots).

Relative age within grade was calculated relative to the same date and was treated as a time-invariant covariate. If the reported birth date was ambiguous over time, the most commonly reported birthdate was assigned. If this strategy did not resolve the discrepancy, the last reported in-range date was selected.

Race/ethnicity was coded from student responses to an ethnicity question (i.e., Hispanic or not) and a race question with a set of “check all that apply” options and the option to give written text responses associated with “other” category. Race/ethnicity was then coded into a mutually exclusive set of categories: Hispanic ethnicity regardless of race; non-Hispanic monoracial categories of white, black, Native Hawaiian/other Pacific Islander, Asian, Native American Indian/Alaska Native; and a single category capturing non-Hispanic multiracial students (explicit claim of multiracial/mixed, multiple claims of individual racial categories). This nominal variable was considered time-invariant, using the most recent responses when reports varied over time.

Guardianship status is time-varying and was constructed from a set of “check all that apply” options (mother, father, stepmother, stepfather, foster mother, foster father, grandmother, grandfather, aunt, uncle, and other) to create two dichotomous indicators representing guardian status where (1) single biological parent and (2) extended family combined with stepparents and foster parents were contrasted with (3) two biological parents.

Students answered TDV questions only if a skip criterion was met (Have you ever DATED someone, including, for example, someone you spent time with or someone you were/are seeing or going out with?). The baseline and follow-up versions of the survey referenced the same recall period for this item. By including dating status as a time-varying covariate, the adjusted TDV variable reflects students who had been in an intimate partner relationship. Students who had never dated prior to or during middle school were omitted from the TDV outcome analysis models.

The administration of the surveys was dependent on the receipt of consent forms and was delayed in some schools. Lag in assessment timing was captured in a set of variables reflecting the number of days relative to a single time point, namely the average assessment date for the sixth-grade fall assessment. The year of the selected anchor date is adjusted for each cohort to reflect the sixth-grade year. To maintain standardization, the month and day remain the same. The first fall semester middle school administration dates ranged from August 2012 to April 2013. The first spring semester middle school data administration ranged from April 2013 to June 2013. Across the 2014–2015 and 2015–2016 school years, middle school fall administration dates ranged from September to March and spring administration ranged from May to October.

For items assessing use of negative conflict styles, students were asked to identify a relationship partner. A dating partner was preferred but if a middle school student reported not having dated during a given assessment period, they were asked to identify a close friend. For this study, “dating partner” is used at the reference category and assigned a value of 0.

***Outlier analysis.*** After covariate adjustment, the data were screened for outliers. The outlier analysis was conducted as an iterative sensitivity analysis, so rather than evaluate outliers within each of the 100 imputed datasets, the grand mean imputed dataset was used to evaluate the optimal settings of the outlier correction routine which was then applied to each of the 100 imputed datasets, separately. The outlier analysis was conducted in three steps: (1) screen for univariate outliers using the Tukey boxplot method46 and the adjusted boxplot method;47 (2) replace univariate outliers with missing data and singly impute these missing data using the principal component auxiliary variables constructed for the missing data treatment as predictors in the imputation model; and (3) plot the treated vs. untreated kernel densities and the treated latent means produced by the modeling strategy described below and compare different parameterizations of the outlier diagnostics. The comparisons in Step 3 were conducted with two goals in mind. First, to minimize the extent to which the outlier correction altered the high-mass portion of the variable’s density. Second, to promote sensible patterns in the latent means (i.e., relatively distinct groupings of Comprehensive and Standard means).

The parameterizations of the two outlier diagnostic routines varied the sensitivity of the detection measured by the number of inner-quartile ranges away from the mean defining an outlier (1.5, 3.0, 4.0). These routines also varied the grouping variables that defined the strata within which outliers were flagged (None, Gender, Gender X Cohort). This setup produced a total of 18 different possible parameterizations to compare [2(Routine) X 3(Sensitivity) X 3(Strata)]. Indicators of the Use of Positive Relationship Styles construct were not outlier-corrected because they were ordinal variables with minimal observed skew. For the three remaining outcomes (TDV Perpetration, TDV Victimization, and Positive Relationship Skills), the sensitivity analysis suggested employing the traditional boxplot approach with a sensitivity of 1.5 inner quartile ranges and grouping by the Gender X Cohort crosstabulation.

***Effects-Coding Parameters.*** In the latent variable models, the loadings and intercept parameters associated with the measurement models were effects-coded to preserve the POMS metric in the latent outcome variables. This nonarbitrary method of scaling imposes the constraint that the loadings should average to 1 and the intercepts should average to 0. The result is a latent mean that is the average of the indicators, and a latent variance that is the average of the reliable variance in the indicators.48 Results of the final measurement models are given in Appendix Table 10.

**Statistical Analysis**

A modeling approach was implemented that is designed to impose parsimony on the means. The freely estimated means were examined in order to iteratively identify means that are described by a single “band” of functionally equal values. Hypothesis tests regarding program effects were based on significant differences between values resulting from constraining means in a given band to a common value, rather than on differences between individual point estimates. This simplification of the model resulted in fewer formal hypothesis tests, balancing the risk of Type I errors (“false positives” that result from tailoring a model to fit a specific sample) and Type II errors (“false negatives” that overlook meaningful program effects). Although not a common approach, it has been used in past research (e.g., Little and Lopez49). Once the constrained bands were identified, the constrained bands of means were then tested for differences between bands. Bands that were not different were re-evaluated and re-assigned to adjacent bands until the bands obtained significant separation. That is, bands that remained different reflect unequivocal differences between the grouped values. When evaluated for direction of effects, the focus was on the preponderance of evidence in support of a treatment effect.

Baseline equivalence across all groups was tested a priori. If the chi-square difference tests revealed significant decrement in fit, a violation of baseline equivalence was inferred. When warranted, baseline equivalence was imposed within gender and/or cohort group and the relative model fit was evaluated. In addition to baseline equivalence, tests of grade level, assessment period within school year, and cohort differences were also incorporated. Specifically, constraints were applied that test the hypothesis that program effects increase over time from the sixth to the eighth grade, reflecting a cumulate effect of participation in the program. Selection of parameter constraints included those testing the hypothesis that means estimated later in the school year (spring assessment) are larger than those estimated earlier (fall assessment), reflecting the effect of exposure over the school year to the classroom-based curricula. Similarly, constraint selection included those testing the hypothesis that program effects are larger for Cohort 4 than for Cohort 3. Since Cohort 4 is one grade level behind Cohort 3 in a given assessment year, cohort differences can be interpreted as the effect of schools implementing the program for a longer period of time, perhaps reflecting familiarity of the schools with the program and a shift in the school-wide culture.

Chi-square difference tests were conducted to evaluate the tenability of the constraints, concluding that a parsimonious set of means successfully described the freely estimated means when the overall chi-square difference test was nonsignificant (*p* > .2). Once a set of constrained means was identified, tests were conducted to determine whether contiguous constrained means were statistically distinct, using post hoc Wald tests (*p* < .01). When adjacent means did not differ, individual means were evaluated for alternative placement or the number of means was decreased.

This approach to hypothesis testing has the potential to detect many discrete means that may be of little practical value when statistical power is high. Too little statistical power may result in few or only one distinct mean, curtailing the opportunity to detect potentially meaningful program effects. In this approach, the constraints were evaluated for evidence of null or counterintuitive effects when the data did not support model constraints showing positive treatment effects.

**Results**

Model fit statistics are included in Appendix Table 11. Program effect size estimates (Cohen’s *d* statistic) are presented in Appendix Table 12. Relative risk reduction for TDV Perpetration, TDV Victimization, and Negative Conflict Styles, calculated as the difference between the DM and SC means divided by the SC mean for each group and time point, are displayed in Appendix Figures 2 through 5.

**Appendix Table 1. Full Sample Demographics and Scores on Teen Dating Violence Perpetration and Victimization, Negative Conflict Resolution Strategies, and Positive Relationship Behaviors**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Full Sample** | **DM** | **SC** |
| Sample size | 2,349 | 1,157 | 1,192 |
| Age, mean (SD) | 11.98 (0.60) | 11.99 (0.61) | 11.97 (0.58) |
| Sex |  |  |  |
|  Males, No. (%) | 1,210 (52) | 586 (51) | 624 (52) |
|  Females, No. (%) | 1,139 (48) | 571 (49) | 568 (48) |
| Cohort |  |  |  |
|  Cohort 3, No. (%) | 1,224 (52) | 615 (53) | 609 (51) |
|  Cohort 4, No. (%) | 1,125 (48) | 542 (47) | 583 (49) |
| Race/Ethnicity |  |  |  |
|  Black, non-Hispanic, No. (%) | 1,282 (55) | 648 (56) | 634 (53) |
|  Hispanic (any race), No. (%) | 648 (28) | 294 (25) | 354 (30) |
|  Multi-racial, non-Hispanic, No. (%) | 163 (7) | 88 (8) | 75 (6) |
|  Asian, non-Hispanic, No. (%) | 145 (6) | 60 (5) | 85 (7) |
|  White, non-Hispanic, No. (%) | 88 (4) | 56 (5) | 32 (3) |
|  Native American/Alaskan Native, No. (%) | 18 (1) | 8 (1) | 10 (1) |
|  Native Hawaiian/other Pacific Islander, No. (%) | 5 (.2) | 3 (.3) | 2 (.2) |

*Note: DM= Dating Matters comprehensive condition; SC= Standard of Care condition. Proportions may sum to greater than 100% due to rounding error. See Appendix for further variable information.*

**Appendix Table 2. Test of Baseline Equivalence on Race, Ethnicity, and Age**

|  |  |  |
| --- | --- | --- |
|  | **Cohort 3** | **Cohort 4** |
|  | **SC** | **DM** |  |  | **SC** | **DM** |  |  |
| **Student Level Characteristic** | ***p or y*** | ***s2*** | ***p or y*** | ***s2*** | **Cox** | **Hedge's *g*** | ***p or y*** | ***s2*** | ***p or y*** | ***s2*** | **Cox** | **Hedge's *g*** |
| **Girls** | (n = 288) | (n = 298) |  |  | (n = 280) | (n = 273) |  |  |
| White | 1.4% |  | 4.7% |  | **0.33** |  | 2.5% |  | 5.9% |  | **0.23** |  |
| Black | 55.2% |  | 53.7% |  | 0.02 |  | 54.6% |  | 58.2% |  | 0.04 |  |
| Pacific Islander | 0.3% |  | 0.3% |  | 0.01 |  | 0.0% |  | 0.4% |  | n/a |  |
| Asian | 5.2% |  | 4.0% |  | **0.07** |  | 7.5% |  | 4.0% |  | **0.17** |  |
| Native American | 1.4% |  | 1.7% |  | **0.05** |  | 0.0% |  | 0.0% |  | n/a |  |
| Mixed | 10.8% |  | 8.7% |  | **0.06** |  | 7.5% |  | 8.1% |  | 0.02 |  |
| Hispanic | 25.7% |  | 26.8% |  | 0.02 |  | 27.9% |  | 23.4% |  | **0.06** |  |
| Age | 11.93 | 0.62 | 11.93 | 0.54 |  | 0.00 | 11.95 | 0.56 | 11.94 | 0.53 |  | 0.00 |
| **Boys** | (n = 321) | (n = 317) |  |  | (n = 303) | (n = 269) |  |  |
| White | 1.6% |  | 4.7% |  | **0.30** |  | 5.3% |  | 4.1% |  | **0.07** |  |
| Black | 52.6% |  | 60.3% |  | **0.08** |  | 50.5% |  | 51.3% |  | 0.01 |  |
| Pacific Islander | 0.0% |  | 0.3% |  | n/a |  | 0.3% |  | 0.0% |  | n/a |  |
| Asian | 5.3% |  | 3.2% |  | **0.14** |  | 10.6% |  | 10.0% |  | 0.01 |  |
| Native American | 0.9% |  | 0.6% |  | **0.10** |  | 1.0% |  | 0.4% |  | **0.26** |  |
| Mixed | 4.4% |  | 4.1% |  | 0.02 |  | 3.0% |  | 10.0% |  | **0.34** |  |
| Hispanic | 35.2% |  | 26.8% |  | **0.10** |  | 29.4% |  | 24.2% |  | **0.07** |  |
| Age | 12.02 | 0.54 | 12.02 | 0.59 |  | 0.00 | 11.96 | 0.57 | 12.07 | 0.67 |  | 0.01 |

*Note:* SC= Standard of Care Condition; DM= Dating Matters comprehensive condition; p= proportion; y= mean;s2= variance. The test statistic shown is the Cox test (proportions) and Hedge's g (means). What Works Clearinghouse's (WWC, 2014) recommendation that Hedge’s g and Cox values of .05–.25 flag characteristics that might be used as covariates in an analysis to correct for potential bias and values above .25 flag characteristics that must be considered when interpreting and generalizing results. Values that exceed this criterion are bolded above.

**Appendix Table 3. Intervention Components by Condition**

|  |  |
| --- | --- |
| *Dating Matters* Comprehensive Prevention Model condition | Standard of Care condition |
| Youth Programs*Dating Matters (*6th grade; 6 sessions)a*Dating Matters* (7th grade; 7 sessions)a*Safe Dates* (8th grade; 10 sessions, poster contest, play)cParent Programs*Parents Matter! for Dating Matters* (6th grade)*b**Dating Matters for Parents* (7th grade)a*Families for Safe* Dates (8th grade)c*i2i: What R U Looking For?* Youth Communications Program a*Dating Matters* Capacity Assessment and Planning Tool a*Dating Matters* Training for Educators a*Dating Matters* Interactive Guide to Informing Policy a*Dating Matters* Guide to Using Indicator Data a | Youth Program*Safe Dates* (8th grade; 10 sessions, poster contest, play)c |

a CDC-developed, evidence-informed interventions. b CDC-adapted, evidence-based parenting program.20,21 c Evidence-based dating violence prevention program.10,16 More information available: [www.cdc.gov/violenceprevention/datingmatters](http://www.cdc.gov/violenceprevention/datingmatters).

**Appendix Table 4. Measure of Teen Dating Violence**

|  |
| --- |
| The following questions ask you about things that may have happened with a boyfriend/girlfriend (past or present) [Follow-up survey version added: *in the last 4 months]*. Fill in the bubbles below that are your best estimates of how often these things have ever happened with someone you were dating. As a guide, use the following scale: |
|  | 1=never (this has never happened in your relationship)2=seldom (this has happened only 1–2 times in your relationship)3=sometimes (this has happened about 3–5 times in your relationship)4=often (this has happened 6 or more times in your relationship) |
| **Subscale** | **Survey Item** |
| Verbal/Emotional Abuse | I did something to make him/her feel jealous. |
|  | I brought up something bad he/she had done in the past. |
|  | I said things just to make him/her angry. |
|  | I spoke to him/her in a hostile or mean tone of voice. |
|  | I insulted him/her with put-downs. |
|  | I ridiculed or made fun of him/her in front of others. |
|  | I kept track of who he/she was with and where he/she was. |
|  | I blamed him/her for the problem. |
|  | I accused him/her of flirting with another girl/guy. |
|  | I threatened to end the relationship. |
| Physical Abuse | I threw something at him/her. |
|  | I kicked, hit, or punched him/her. |
|  | I slapped him/her or pulled his/her hair. |
|  | I pushed, shoved, or shook him/her. |
| Relational Abuse | I tried to turn his/her friends against him/her. |
|  | I said things to his/her friends about him/her to turn them against him/her. |
|  | I spread rumors about him/her. |
| Sexual Abuse | I touched him/her sexually when he/she didn’t want me to. |
|  | I forced him/her to have sex when he/she didn’t want to. |
|  | I threatened him/her in an attempt to have sex with him/her. |
|  | I kissed him/her when he/she didn’t want me to. |
| Threatening Behavior | I destroyed or threatened to destroy something he/she valued. |
|  | I deliberately tried to frighten him/her. |
|  | I threatened to hurt him/her. |
|  | I threatened to hit him/her or throw something at him/her. |
|  | I threatened him/her with a knife or gun (including waving or pointing a knife). |
| Severe Physical Abuse | I choked him/her. |
|  | I used a knife or fired a gun. |
|  | I scratched him/her and/or bent his/her fingers. |
|  | I burned him/her. |
|  | I bit him/her. |
| Note: Items listed here reflect wording used for perpetration. Each item was repeated, using wording reflecting victimization. |

**Appendix Table 5. Measure of Negative Conflict Resolution Strategies**

|  |
| --- |
| The following questions refer to times when you and the person you are dating have disagreements. We want you to answer these questions about a dating partner if you have one now or have had one in the last 4 months. But if you do not have a current or recent (in the last 4 months) boyfriend or girlfriend, please think about a close friend, either a boy or a girl. How often do YOU use these styles… |
|  | 1=never2=almost never3=sometimes4=often5=always |
| **Subscale** | **Survey Item** |
| Compliance | Not being willing to stick up for myself |
|  | Reaching a limit, shutting down, and refusing to talk any further |
|  | Not defending my position |
|  | Giving in with little attempt to present my side of the issue |
| Conflict Engagement | Launching personal attacks |
|  | Exploding and getting out of control |
|  | Getting carried away and saying things that aren’t meant |
|  | Throwing insults and digs |
| Withdrawal | Remaining silent for long periods of time |
|  | Being too compliant |
|  | Tuning the other person out |
|  | Withdrawing, acting distant, and not interested |

Note: These items were preceded with the following prompt:

*Please remember that by “dating,” we mean spending time with someone you are seeing or going out with. Examples of this might include hanging out at the mall, in the neighborhood, or at home or going somewhere together like the movies, a game, or a party. It doesn't have to be a formal date or something you planned in advance and it may be with a small group. The term "date" includes both one-time dates and time together as part of long-term relationships.*

**Appendix Table 6. Measure of Positive Relationship Skills**

|  |
| --- |
| Please indicate how often each of the following statements is true currently or was true in your most recent dating relationship [follow-up survey version added “in the last 4 months”]. |
| 1=never2=sometimes3=usually4=always |
| My boyfriend/girlfriend is/was honest and truthful with me. |
| My boyfriend/girlfriend and I are/were good at working out our differences. |
| When I have a serious disagreement with my boyfriend/girlfriend, we discuss(ed) it respectfully. |
| My boyfriend/girlfriend and I work(ed) as a team. |

**Appendix Table 7a. Means of Outcome Variables Using Item Averages (Original Metric)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Grade 6*** | ***Grade 7*** | ***Grade 8*** |
|  | ***Fall*** | ***Spring*** | ***Fall*** | ***Spring*** | ***Fall*** | ***Spring*** |
| **Group** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Teen Dating Violence Perpetration (1=”never”, 4=”often”) |
| Cohort 3 Females SC | 1.22 | 0.28 | 1.30 | 0.36 | 1.22 | 0.28 | 1.26 | 0.36 | 1.23 | 0.32 | 1.25 | 0.33 |
| Cohort 3 Females DM | 1.21 | 0.32 | 1.25 | 0.33 | 1.20 | 0.29 | 1.22 | 0.34 | 1.19 | 0.26 | 1.22 | 0.32 |
| Cohort 3 Males SC | 1.20 | 0.29 | 1.26 | 0.36 | 1.16 | 0.24 | 1.21 | 0.36 | 1.17 | 0.25 | 1.22 | 0.35 |
| Cohort 3 Males DM | 1.22 | 0.34 | 1.25 | 0.38 | 1.16 | 0.24 | 1.22 | 0.38 | 1.18 | 0.29 | 1.20 | 0.32 |
| Cohort 4 Females SC | 1.19 | 0.24 | 1.25 | 0.34 | 1.19 | 0.26 | 1.23 | 0.33 | 1.18 | 0.29 | 1.19 | 0.26 |
| Cohort 4 Females DM | 1.24 | 0.32 | 1.22 | 0.30 | 1.19 | 0.29 | 1.20 | 0.34 | 1.19 | 0.28 | 1.23 | 0.35 |
| Cohort 4 Males SC | 1.17 | 0.26 | 1.21 | 0.31 | 1.15 | 0.24 | 1.17 | 0.28 | 1.13 | 0.17 | 1.18 | 0.28 |
| Cohort 4 Males DM | 1.20 | 0.30 | 1.22 | 0.38 | 1.17 | 0.29 | 1.17 | 0.31 | 1.16 | 0.29 | 1.23 | 0.33 |
| Teen Dating Violence Victimization (1=”never”, 4=”often”) |
| Cohort 3 Females SC | 1.22 | 0.27 | 1.30 | 0.36 | 1.22 | 0.28 | 1.25 | 0.35 | 1.22 | 0.31 | 1.26 | 0.37 |
| Cohort 3 Females DM | 1.20 | 0.29 | 1.26 | 0.35 | 1.20 | 0.28 | 1.23 | 0.35 | 1.19 | 0.27 | 1.22 | 0.34 |
| Cohort 3 Males SC | 1.23 | 0.30 | 1.27 | 0.37 | 1.19 | 0.27 | 1.22 | 0.33 | 1.20 | 0.29 | 1.24 | 0.36 |
| Cohort 3 Males DM | 1.25 | 0.34 | 1.26 | 0.37 | 1.18 | 0.26 | 1.25 | 0.39 | 1.21 | 0.34 | 1.21 | 0.33 |
| Cohort 4 Females SC | 1.19 | 0.25 | 1.24 | 0.30 | 1.19 | 0.25 | 1.21 | 0.27 | 1.19 | 0.31 | 1.20 | 0.27 |
| Cohort 4 Females DM | 1.23 | 0.33 | 1.23 | 0.30 | 1.19 | 0.28 | 1.22 | 0.36 | 1.20 | 0.31 | 1.23 | 0.35 |
| Cohort 4 Males SC | 1.19 | 0.28 | 1.25 | 0.35 | 1.17 | 0.25 | 1.18 | 0.30 | 1.15 | 0.20 | 1.18 | 0.26 |
| Cohort 4 Males DM | 1.23 | 0.30 | 1.24 | 0.38 | 1.19 | 0.29 | 1.20 | 0.34 | 1.18 | 0.32 | 1.23 | 0.32 |
| Negative Conflict Resolution (1=”never”, 5=”always”) |
| Cohort 3 Females SC | 1.99 | 0.68 | 1.98 | 0.68 | 1.91 | 0.68 | 2.01 | 0.74 | 2.03 | 0.74 | 2.06 | 0.74 |
| Cohort 3 Females DM | 1.85 | 0.62 | 2.00 | 0.71 | 1.94 | 0.71 | 1.99 | 0.74 | 1.95 | 0.72 | 1.98 | 0.73 |
| Cohort 3 Males SC | 1.87 | 0.65 | 1.87 | 0.68 | 1.77 | 0.65 | 1.78 | 0.63 | 1.77 | 0.68 | 1.81 | 0.70 |
| Cohort 3 Males DM | 1.83 | 0.63 | 1.91 | 0.72 | 1.87 | 0.68 | 1.86 | 0.69 | 1.85 | 0.69 | 1.89 | 0.71 |
| Cohort 4 Females SC | 1.89 | 0.67 | 1.94 | 0.72 | 1.93 | 0.72 | 1.99 | 0.72 | 1.90 | 0.77 | 2.07 | 0.80 |
| Cohort 4 Females DM | 1.93 | 0.67 | 1.92 | 0.68 | 1.86 | 0.70 | 1.87 | 0.68 | 1.83 | 0.66 | 1.98 | 0.72 |
| Cohort 4 Males SC | 1.72 | 0.57 | 1.76 | 0.61 | 1.75 | 0.64 | 1.81 | 0.66 | 1.77 | 0.75 | 1.93 | 0.74 |
| Cohort 4 Males DM | 1.75 | 0.59 | 1.77 | 0.64 | 1.70 | 0.59 | 1.71 | 0.63 | 1.76 | 0.68 | 1.87 | 0.68 |
| Positive Relationship Skills (1=”never”, 4=”always”) |
| Cohort 3 Females SC | 2.70 | 0.98 | 2.75 | 0.96 | 2.85 | 0.91 | 2.93 | 0.86 | 2.97 | 0.80 | 2.90 | 0.90 |
| Cohort 3 Females DM | 2.80 | 1.03 | 2.80 | 0.94 | 2.89 | 0.92 | 2.92 | 0.93 | 3.02 | 0.87 | 2.98 | 0.87 |
| Cohort 3 Males SC | 2.82 | 0.99 | 2.79 | 0.95 | 2.80 | 0.94 | 2.94 | 0.90 | 2.97 | 0.89 | 2.94 | 0.89 |
| Cohort 3 Males DM | 2.90 | 0.98 | 2.87 | 0.94 | 2.93 | 0.92 | 3.05 | 0.87 | 2.96 | 0.91 | 2.91 | 0.88 |
| Cohort 4 Females SC | 2.69 | 1.00 | 2.71 | 0.96 | 2.68 | 0.96 | 2.92 | 0.90 | 3.04 | 0.92 | 2.98 | 0.88 |
| Cohort 4 Females DM | 2.84 | 1.01 | 2.84 | 0.96 | 2.76 | 0.98 | 2.99 | 0.88 | 2.95 | 0.90 | 2.97 | 0.82 |
| Cohort 4 Males SC | 2.90 | 0.97 | 2.90 | 0.95 | 2.95 | 0.97 | 3.03 | 0.92 | 3.02 | 0.88 | 2.96 | 0.91 |
| Cohort 4 Males DM | 2.89 | 0.99 | 2.86 | 0.96 | 2.78 | 0.92 | 2.94 | 0.87 | 3.00 | 0.90 | 2.96 | 0.88 |

**Appendix Table 7B. Means of Outcome Latent Variables (POMS Metric)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Grade 6*** | ***Grade 7*** | ***Grade 8*** |
|  | ***Fall*** | ***Spring*** | ***Fall*** | ***Spring*** | ***Fall*** | ***Spring*** |
| **Group** | ***M*** | ***SE*** | ***M*** | ***SE*** | ***M*** | ***SE*** | ***M*** | ***SE*** | ***M*** | ***SE*** | ***M*** | ***SE*** |
| Teen Dating Violence Perpetration (0=all items “never”, 100=all items “often”) |
| Cohort 3 Females SC | 5.42 | 0.44 | 7.70 | 0.37 | 4.37 | 0.18 | 5.22 | 0.32 | 4.31 | 0.17 | 5.68 | 0.29 |
| Cohort 3 Females DM | 5.15 | 0.36 | 7.03 | 0.32 | 4.45 | 0.17 | 5.03 | 0.28 | 4.22 | 0.16 | 5.48 | 0.28 |
| Cohort 3 Males SC | 5.11 | 0.24 | 7.22 | 0.33 | 4.48 | 0.17 | 4.82 | 0.23 | 4.31 | 0.16 | 5.40 | 0.28 |
| Cohort 3 Males DM | 5.00 | 0.27 | 6.84 | 0.33 | 4.38 | 0.14 | 4.87 | 0.23 | 4.31 | 0.16 | 5.18 | 0.25 |
| Cohort 4 Females SC | 5.39 | 0.25 | 6.77 | 0.32 | 4.25 | 0.17 | 5.33 | 0.29 | 4.52 | 0.17 | 5.59 | 0.27 |
| Cohort 4 Females DM | 5.34 | 0.35 | 6.32 | 0.30 | 4.26 | 0.17 | 4.90 | 0.27 | 4.36 | 0.18 | 5.44 | 0.29 |
| Cohort 4 Males SC | 5.31 | 0.20 | 6.94 | 0.32 | 4.41 | 0.16 | 4.76 | 0.21 | 4.49 | 0.14 | 5.36 | 0.25 |
| Cohort 4 Males DM | 5.11 | 0.20 | 6.12 | 0.34 | 4.35 | 0.17 | 4.33 | 0.23 | 4.27 | 0.16 | 5.55 | 0.29 |
| Teen Dating Violence Victimization (0=all items “never”, 100=all items “often”) |
| Cohort 3 Females SC | 5.79 | 0.31 | 8.37 | 0.42 | 5.48 | 0.25 | 6.25 | 0.32 | 4.89 | 0.24 | 6.17 | 0.30 |
| Cohort 3 Females DM | 5.21 | 0.28 | 7.71 | 0.40 | 5.37 | 0.22 | 5.83 | 0.28 | 4.64 | 0.22 | 5.59 | 0.27 |
| Cohort 3 Males SC | 6.15 | 0.32 | 7.87 | 0.41 | 5.33 | 0.22 | 5.84 | 0.29 | 4.97 | 0.23 | 5.94 | 0.28 |
| Cohort 3 Males DM | 5.94 | 0.35 | 7.52 | 0.40 | 5.31 | 0.20 | 6.09 | 0.29 | 5.01 | 0.23 | 5.39 | 0.27 |
| Cohort 4 Females SC | 5.97 | 0.24 | 7.45 | 0.33 | 5.17 | 0.23 | 6.16 | 0.28 | 5.24 | 0.21 | 6.01 | 0.28 |
| Cohort 4 Females DM | 5.98 | 0.28 | 6.86 | 0.35 | 5.20 | 0.23 | 5.88 | 0.29 | 4.96 | 0.24 | 5.75 | 0.29 |
| Cohort 4 Males SC | 6.05 | 0.25 | 7.70 | 0.37 | 5.21 | 0.22 | 5.56 | 0.26 | 5.26 | 0.21 | 5.62 | 0.26 |
| Cohort 4 Males DM | 6.11 | 0.26 | 6.88 | 0.39 | 5.32 | 0.23 | 5.33 | 0.29 | 4.82 | 0.22 | 5.85 | 0.32 |
| Negative Conflict Resolution (0=all items “never”, 100=all items “always”) |
| Cohort 3 Females SC | 28.48 | 1.04 | 29.03 | 1.11 | 27.23 | 1.03 | 29.92 | 1.10 | 29.94 | 1.08 | 31.13 | 1.10 |
| Cohort 3 Females DM | 25.36 | 0.98 | 29.64 | 1.09 | 28.69 | 1.01 | 28.87 | 1.07 | 29.01 | 1.06 | 29.39 | 1.09 |
| Cohort 3 Males SC | 26.13 | 0.96 | 27.17 | 1.05 | 24.29 | 0.92 | 24.40 | 0.91 | 24.07 | 0.95 | 25.12 | 1.02 |
| Cohort 3 Males DM | 25.27 | 0.93 | 28.11 | 1.05 | 27.07 | 0.95 | 26.37 | 0.97 | 26.50 | 1.00 | 27.54 | 1.04 |
| Cohort 4 Females SC | 26.66 | 1.04 | 28.89 | 1.08 | 27.71 | 1.07 | 29.38 | 1.09 | 28.50 | 1.15 | 31.60 | 1.23 |
| Cohort 4 Females DM | 27.03 | 1.05 | 27.19 | 1.03 | 26.78 | 1.07 | 26.29 | 1.07 | 26.80 | 1.05 | 28.96 | 1.15 |
| Cohort 4 Males SC | 23.05 | 0.88 | 24.23 | 0.94 | 23.39 | 0.93 | 25.23 | 1.00 | 25.00 | 1.08 | 27.77 | 1.06 |
| Cohort 4 Males DM | 23.25 | 0.90 | 24.12 | 1.02 | 23.27 | 0.94 | 23.18 | 1.01 | 25.44 | 1.08 | 26.48 | 1.10 |
| Positive Relationship Skills (0=all items “never”, 100=all items “always”) |
| Cohort 3 Females SC | 59.42 | 2.02 | 60.83 | 2.09 | 63.55 | 1.99 | 61.69 | 1.91 | 62.51 | 1.90 | 60.66 | 2.04 |
| Cohort 3 Females DM | 61.95 | 2.10 | 61.56 | 2.12 | 64.71 | 1.92 | 60.73 | 1.92 | 63.64 | 1.95 | 63.50 | 2.00 |
| Cohort 3 Males SC | 62.63 | 1.93 | 61.59 | 1.92 | 62.44 | 1.93 | 62.73 | 1.84 | 62.90 | 1.87 | 62.65 | 1.93 |
| Cohort 3 Males DM | 65.02 | 1.99 | 63.69 | 1.98 | 65.58 | 1.87 | 65.36 | 1.81 | 60.97 | 1.93 | 60.72 | 1.99 |
| Cohort 4 Females SC | 57.25 | 2.06 | 58.90 | 2.05 | 57.83 | 2.05 | 61.60 | 1.85 | 63.95 | 1.96 | 62.61 | 2.01 |
| Cohort 4 Females DM | 62.17 | 2.18 | 62.36 | 2.09 | 60.68 | 2.20 | 62.45 | 1.98 | 61.02 | 2.13 | 62.05 | 1.98 |
| Cohort 4 Males SC | 64.43 | 2.01 | 64.38 | 1.97 | 66.23 | 2.02 | 64.78 | 1.94 | 63.67 | 1.84 | 62.91 | 2.05 |
| Cohort 4 Males DM | 64.31 | 2.16 | 63.89 | 2.12 | 61.32 | 2.13 | 61.09 | 1.90 | 62.44 | 2.03 | 61.95 | 2.12 |

Note: Item parcel indicators of the latent variables were covariate-adjusted and corrected for outliers (see pages 5-7). The values reported above are the unconstrained estimates used as the baseline model against which constraint selections were evaluated.

**Appendix Table 8. Cronbach’s Alpha Reliability Coefficients**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Outcome** | **Fall 6th** | **Spring 6th** | **Fall 7th** | **Spring 7th** | **Fall 8th** | **Spring 8th** |
| Teen Dating Violence Perpetration | 0.90 | 0.92 | 0.89 | 0.93 | 0.91 | 0.92 |
| Teen Dating Violence Victimization | 0.90 | 0.91 | 0.89 | 0.93 | 0.92 | 0.92 |
| Use of Negative Conflict Resolution Strategies | 0.76 | 0.79 | 0.82 | 0.82 | 0.85 | 0.85 |
| Positive Relationship Skills | 0.85 | 0.81 | 0.87 | 0.84 | 0.88 | 0.84 |

**Appendix Table 9. Conversion of Original Response Anchors to POMS Scores**

|  | **Original Response Categories** | **Equivalent POMS Score** |
| --- | --- | --- |
| TDV |  |  |
| never | 1 | 0 |
| seldom | 2 | 33 |
| sometimes | 3 | 66 |
| often | 4 | 100 |
| Use of Negative Conflict Styles |  |  |
| never | 1 | 0 |
| almost never | 2 | 25 |
| sometimes | 3 | 50 |
| often | 4 | 75 |
| always | 5 | 100 |
| Positive Relationship Skills |  |  |
| never | 1 | 0 |
| sometimes | 2 | 33 |
| usually | 3 | 66 |
| always | 4 | 100 |

**Appendix Table 10. Unstandardized Measurement Model Results**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **TDV Perpetration** |  | **TDV Victimization** |  | **Negative Conflict Styles** |  | **Positive Relationship Skills** |  |
| **Parcels** | **Loading** | **SE** |  | **Loading** | **SE** |  | **Loading** | **SE** |  | **Loading** | **SE** |  |
| Parcel 1 | 1.59 | 0.05 |  | 1.50 | 0.04 |  | 0.99 | 0.01 |  | 0.97 | 0.01 |  |
| Parcel 2 | 0.64 | 0.04 |  | 1.11 | 0.04 |  | 0.96 | 0.01 |  | 1.02 | 0.01 |  |
| Parcel 3 | 0.78 | 0.05 |  | 0.39 | 0.04 |  | 1.05 | 0.01 |  | 1.01 | 0.01 |  |
| Fit Statistics |  |
| Chi-Square | 776.23 |  | 756.12 |  | 583.71 |  | 237.66 |  |
| df | 788 |  | 788 |  | 788 |  | 788 |  |
| *p*-value | 0.611 |  | 0.787 |  | 1.000 |  | 1.000 |  |
| RMSEA | 0.00 |  | 0.00 |  | 0.00 |  | 0.00 |  |
| SRMR | 0.08 |  | 0.08 |  | 0.04 |  | 0.03 |  |
| CFI | 1.00 |  | 1.00 |  | 1.00 |  | 1.00 |  |
| TLI | 1.01 |  | 1.02 |  | 1.04 |  | 1.12 |  |

**Appendix Table 11. Results of Model Tests**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Unconstrained** | **Constrained** | **Difference** |
|  | **Chi-square** | **df** | **RMSEA** | **SRMR** | **Chi-square** | **df** | **RMSEA** | **SRMR** | **Chi-square** | **df** | **P** |
| TDV-P | 776.23 | 788 | 0.00 | 0.08 | 803.76 | 830 | 0.00 | 0.08 | 27.53 | 42 | 0.958 |
| TDV-V | 776.23 | 788 | 0.00 | 0.08 | 825.81 | 831 | 0.00 | 0.08 | 49.58 | 43 | 0.227 |
| NCRS | 583.71 | 788 | 0.00 | 0.04 | 630.23 | 831 | 0.00 | 0.04 | 46.52 | 43 | 0.330 |
| UPRS | 237.66 | 788 | 0.00 | 0.03 | 256.50 | 830 | 0.00 | 0.03 | 18.85 | 42 | 0.999 |
|  |  |  |  |  |
|  |  |  |  | **Pairwise Diff Tests (adjacent constrained means)** |  |
| **TDV-P** |  |  |  |  |  |  | **95% CI** |  |  |  |  |
|  |  | **Mean** | **SE** |  **Pair** | **Diff** | **SE** | **Lower** | **Upper** | **P** |  |  |   |
|  |  *A:* | 4.27 | 0.09 | *A vs B:* | 0.26 | 0.08 | 0.10 | 0.41  | 0.001 |  |  |   |
|  |  *B:* | 4.53 | 0.08  | *B vs C:* | 0.66 | 0.11 | 0.44 | 0.88  | 0.000 |  |  |   |
|  |  *C:* | 5.19 | 0.11  | *C vs D:* | 0.57 | 0.17 | 0.23 | 0.91  | 0.001 |  |  |   |
|  |  *D:* | 5.76 | 0.16 | *D vs E:* | 1.22 | 0.20 | 0.82 | 1.61  | 0.000 |  |  |   |
|  |  *E:* | 6.98 | 0.17 |  |   |  |   |   |   |  |  |   |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **TDV-V** |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  *A:* | 5.02 | 0.09  | *A vs B* | 0.34 | 0.10 | 0.15 | 0.53  | 0.000 |  |  |  |
|  |  *B:* | 5.36 | 0.09 | *B vs C* | 0.56 | 0.11 | 0.34 | 0.76  | 0.000 |  |  |  |
|  |  *C:* | 5.92 | 0.09  | *C vs D* | 1.01 | 0.24 | 0.54 | 1.48  | 0.000 |  |  |  |
|  |  *D:* | 6.93 | 0.23 | *D vs E* | 0.86 | 0.25 | 0.37 | 1.35  | 0.001 |  |  |  |
|  |  *E:* | 7.79 | 0.23  |  |  |   |   |  |   |   |   |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **NCRS** |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  *A:* | 23.68 | 0.48 | *A vs B* | 1.88 | 0.49 | 0.92 | 2.84  | 0.000 |  |  |  |
|  |  *B:* | 25.55 | 0.32  | *B vs C* | 1.38 | 0.35 | 0.70 | 2.06  | 0.000 |  |  |  |
|  |  *C:* | 26.93 | 0.31 | *C vs D* | 1.40 | 0.43 | 0.56 | 2.23  | 0.001 |  |  |  |
|  |  *D:* | 28.33 | 0.43  | *D vs E* | 2.33 | 0.63 | 1.09 | 3.57  | 0.000 |  |  |  |
|  |  *E:* | 30.66 | 0.66  |  |  |   |   |   |   |   |   |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **UPRS** |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  A: | 62.38 | 0.43 |  |  |  |  |  |  |  |  |  |

Note: Each outcome was independently evaluated. TDV-P signifies teen dating violence perpetration, TDV-V signifies teen dating violence victimization, UNCS signifies use of negative conflict styles, UPRS signifies use of positive relationship skills. The letters A through E signify the means resulting from the model constraints. Because the final model of UPRS did not reveal multiple means constraints, no pairwise difference tests were conducted. Confidence intervals are given for the difference tests, on which the conclusions are based, rather than the individual constrained means.

**Appendix Table 12. Program Effect Size Estimates (Cohen’s *d* statistic)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TDV-P | Cohort 3 Girls | Cohort 4 Girls | Cohort 3 Boys | Cohort 4 Boys |
| Grade 6 Spring | 0.00 | -0.02 | 0.00 | -0.02 |
| Grade 7 Fall | -0.01 | -0.01 | -0.01 | -0.01 |
| Grade 7 Spring | -0.02 | -0.02 | -0.02 | -0.03 |
| Grade 8 Fall | -0.01 | -0.01 | -0.01 | -0.01 |
| Grade 8 Spring | -0.01 | -0.01 | -0.01 | 0.00 |
| TDV-V |  |  |  |  |
| Grade 6 Spring | -0.01 | -0.01 | -0.01 | -0.01 |
| Grade 7 Fall | -0.02 | -0.01 | -0.02 | -0.01 |
| Grade 7 Spring | -0.01 | -0.02 | -0.01 | -0.01 |
| Grade 8 Fall | -0.01 | -0.01 | -0.01 | -0.01 |
| Grade 8 Spring | -0.01 | -0.01 | -0.01 | -0.01 |
| NCRS |  |  |  |  |
| Grade 6 Spring | 0.00 | -0.01 | 0.00 | 0.00 |
| Grade 7 Fall | -0.01 | -0.01 | 0.00 | -0.01 |
| Grade 7 Spring | -0.01 | -0.01 | 0.00 | -0.01 |
| Grade 8 Fall | -0.01 | -0.01 | 0.00 | -0.01 |
| Grade 8 Spring | -0.01 | -0.01 | 0.00 | -0.01 |

Note: There were no program effects for Positive Relationship Skills (PRS) so effect size estimates are not shown.

**Appendix Figure 1. Data Assessment Design**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **6th Grade** | **7th Grade** | **8th Grade** |
|  | **Fall** | **Spring** | **Fall** | **Spring** | **Fall** | **Spring** |
| Cohort 1 |  |  |  |  | Y1 | Y1 |
| Cohort 2 |  |  | Y1 | Y1 | Y2 | Y2 |
| Cohort 3 | **Y1** | **Y1** | **Y2** | **Y2** | **Y3** | **Y3** |
| Cohort 4 | **Y2** | **Y2** | **Y3** | **Y3** | **Y4** | **Y4** |
| Cohort 5 | Y3 | Y3 | Y4 | Y4 |  |  |
|  |

Note: The letter Y in the figure represents the calendar year of school-level assessment. Y1 = 2012–2013; Y2 = 2013–2014; Y3 = 2014–2015; Y4 = 2015–2016. All six time points were included in the data collection design for Cohorts 3 and 4 (bolded). All other cohorts had two or more time points that were missing by design.

**Appendix Figure 2. Relative Risk Reduction for DM vs. SC by cohort, gender, and time periods: Teen Dating Violence Perpetration**

Note. Relative risk reduction represents the percent reduction in scores on the measure of teen dating violence perpetration for the Dating Matters comprehensive condition relative to the standard-of-care condition.

**Appendix Figure 3. Relative Risk Reduction Relative Risk Reduction for DM vs. SC by cohort, gender, and time periods: Teen Dating Violence Victimization**

Note. Relative risk reduction represents the percent reduction in scores on the measure of teen dating violence victimization for the Dating Matters comprehensive condition relative to the standard-of-care condition.

**Appendix Figure 4. Relative Risk Reduction Relative Risk Reduction for DM vs. SC by cohort, gender, and time periods: Negative Conflict Resolution**

Note. Relative risk reduction represents the percent reduction in scores on the measure of negative conflict resolution for the Dating Matters comprehensive condition relative to the standard-of-care condition.

**Appendix Figure 5. Percent relative risk reduction (M, range) for DM vs. SC across cohorts and time periods**

Note. Relative risk reduction represents the percent reduction in scores on measures of teen dating violence, negative conflict resolution, and positive relationship behaviors for the Dating Matters comprehensive condition relative to the standard-of-care condition. The numbers within the circles represent the average risk reduction for that outcome across the 4 groups (cohort x sex), and the space between the diamonds represent the range of relative risk reduction on that outcome across the four groups.

**Appendix Figure 6. Constrained Means Across Time by Sex and Cohort – Positive Relationship Skills**

Note. SC= Standard of Care condition. DM= Dating Matters condition. Sample sizes (Ns) for each condition within each group are reported next to the condition label of the respective line in each figure. Percent of Maximum Score (POMS) refers to the maximum possible score given the number of items and response categories in a scale, rather than the maximum observed score. Non-overlapping lines represent significant group differences. Standard errors, confidence intervals, and statistical significance for each estimated mean value is reported in Appendix Table 11.

**References**

1. Black BM, Chido LM, Preble KM, et al. Violence exposure and teen dating violence among african american youth. *J Interpers Violence.* 2015;30(12):2174-2195.

2. Cunradi CB, Caetano R, Clark C, Schafer J. Neighborhood poverty as a predictor of intimate partner violence among White, Black, and Hispanic couples in the United States: A multilevel analysis. *Ann Epidemiol.* 2000;10(5):297-308.

3. Gorman-Smith D, Tolan P. The role of exposure to community violence and developmental problems among inner-city youth. *Dev Psychopathol.* 1998;10(1):101-116.

4. Halliday-Boykins CA, Graham S. At both ends of the gun: Testing the relationship between community violence exposure and youth violent behavior. *J Abnorm Child Psychol.* 2001;29(5):383-402.

5. Johnson RM, Parker EM, Rinehart J, Nail J, Rothman EF. Neighborhood factors and dating violence among youth. *Am J Prev Med.* 2015;49(3):458-466.

6. Sampson RJ, Lauritsen JL. Violent victimization and offending: Individual-, situational-, and community-level risk factors. In: Reiss AJ, Roth JA, ed. *Understanding and preventing violence, volume 3, social influences*. Washington, D.C.: National Academy Press; 1994:1-114.

7. What Works Clearinghouse. *Procedures and standards handbook (Version 3).* 2014.

8. Teten Tharp A. Dating Matters: The next generation of teen dating violence prevention. *Prev Sci.* 2012;13:398-401.

9. Teten Tharp A, Burton T, Freire K, et al. Dating Matters™: Strategies to Promote Healthy Teen Relationships. *J Womens Health.* 2011;20:1-5.

10. Foshee VA, Bauman KE, Ennett ST, Linder GF, Benefield T, Suchindran C. Assessing the long-term effects of the Safe Dates program and a booster in preventing and reducing adolescent dating violence victimization and perpetration. *Am J Public Health.* 2004;94(4):619-624.

11. David-Ferdon C, Simon TR. *Preventing youth violence: Opportunities for action.* Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention; 2014.

12. Matjasko JL, Vivolo-Kantor AM, Massetti GM, Holland KM, Holt MK, Dela Cruz J. A systematic meta-review of evaluations of youth violence prevention programs: Common and divergent findings from 25 years of meta-analyses and systematic reviews. *Aggress Violent Behav.* 2012;17(6):540-552.

13. Wolfe DA, Crooks CV, Jaffe P, et al. A school-based program to prevent adolescent dating violence: A cluster randomized trial. *Arch Pediatr Adolesc Med.* 2009;163:692-699.

14. Andrews JA, Foster S, Capaldi D, Hops H. Adolescent and family predictors of physical aggression, communication, and satisfaction in young adult couples: A prospective analysis. *J Consult Clin Psychol*. 2000;68:195-208.

15. Foshee, VA.; Matthew, RA. Cambridge Handbook of Violent Behavior. Cambridge University; New York, New York: 2007. Adolescent Dating Abuse Perpetration: A Review of Findings, Methodology Limitations and Suggestions for Future Research.; p. 431-449.

16. Foshee VA, McNaughton Reyes HL, Ennett ST, Cance JD, Bauman KE, Bowling M. Assessing the effects of Families for Safe Dates, a family-based teen dating abuse prevention program. *J Adoles Health.* 2012;51(4):349-356.

17. Lavoie F, Hebert M, Tremblay R, Vitaro F, Vezina L, McDuff P. History of family dysfunction and perpetration of dating violence by adolescent boys: A longitudinal study. *J Adol Health*. 2002;31:375-383.

18. Tschann JM, Pasch LA, Flores E, VanOss Marin B, Baisch EM, Wibbelsman CJ. Nonviolent aspects of interparental conflict and dating violence among adolescents. *J Fam Issues*. 2009;30:295-319.

19. Tyler KA, Brownridge DA, Melander LA. The effect of poor parenting on male and female dating violence perpetration and victimization. *Violence Vict.* 2001;26:218-230.

20 Dittus P, Miller KS, Kotchick BA, Forehand R. Why parents matter! The conceptual basis for a community-based HIV prevention program for the parents of African American youth. *J Child Fam Stud*. 2004;13:5–20.

21. Forehand R, Armistead L, Long N, et al. Efficacy of a family based, youth sexual risk program for parents of African American pre-adolescents: A randomized controlled design. *Arch Pediatr Adolesc Med*. 2007;161:1123–1129.

22. Lang KM, Little TD. Principled missing data treatments. *Prev Sci.* 2018;19:284-294.

23. Enders CK. *Applied missing data analysis.* New York: Guilford Press; 2010.

24. Howard WJ, Rhemtulla M, Little TD. Using principal components as auxiliary variables in missing data estimation. *Multivariate Behav Res.* 2015;50(3):285-299.

25. Lang KM, Little TD, PcAux Development Team. PcAux: Automatically extract auxiliary features for simple, principled missing data analysis (R package version 0.0.0.9004). 2017; <http://github.com/PcAux-Package/PcAux/>.

26. Team RC. *R: A language and environment for statistical computing [Computer software manual]. .* Vienna, Austria: Retrieved from http://www.R-project.org/; 2017.

27. Little TD, Jorgensen TD, Lang KM, Moore EW. On the joys of missing data. *J Pediatr Psychol.* 2014;39(2):151-162.

28. van Buuren S. *Flexible imputation of missing data.* Boca Raton, FL: CRC Press; 2012.

29. Van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. *J Stat Comput Simul.* 2006;76(12):1049-1064.

30. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning (2nd ed.).* New York, NY: Springer; 2009.

31. Rubin DB. *Multiple imputation for nonresponse in surveys.* New York, NY: John Wiley & Sons; 1987.

32. von Hippel PT. How to impute interactions, squares, and other transformed variables. *Sociol Methodol.* 2009;39(1):265-291.

33. Little TD. *Longitudinal structural equation modeling.* New York: Guilford Press; 2013.

34. Little TD, Cunningham WA, Shahar G, Widaman KF. To parcel or not to parcel: Exploring the question, weighing the merits. *Struct Equ Model.* 2002;9(2):151-173.

35. Little TD, Rhemtulla M, Gibson K, Schoemann AM. Why the items versus parcels controversy needn't be one. *Psychol Methods.* 2013;18(3):285-300.

36. Wolfe DA, Wekerle C, Scott K, Straatman AL, Grasley C, Reitzel-Jaffe D. Dating violence prevention with at-risk youth: a controlled outcome evaluation. *J Consult Clinical Psychol.* 2003;71:279.

37. Wolfe DA, Scott K, Reitzel-Jaffe D, Wekerle C, Grasley C, Straatman AL. Development and validation of the conflict in adolescent dating relationships inventory. *Psychol Assess.* 2001;13:277.

38. Teitelman AM, Ratcliffe SJ, Morales-Aleman MM, Sullivan CM. Sexual relationship power, intimate partner violence, and condom use among minority urban girls. *J Interpers Violence.* 2008;23(12):1694-1712.

39. Hokoda A, Ramos-Lira L, Celaya P, et al. Reliability of translated measures assessing dating violence among Mexican adolescents. *Violence Vict.* 2006;21(1):117-127.

40. Schiff M, Zeira A. Dating violence and sexual risk behaviors in a sample of at-risk Israeli youth. *Child Abuse Negl.* 2005;29(11):1249-1263.

41. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *J Pers Assess.* 2010;92(6):544-559.

42. Gomez R, McLaren S. The center for epidemiologic studies depression scale: support for a bifactor model with a dominant general factor and a specific factor for positive affect. *Assessment.* 2015;22(3):351-360.

43. Lance CE. Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interactions. *Appl Psychol Meas.* 1988;12(2):163-175.

44. Cohen J, Cohen P, West SG, Aiken LS. *Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.).* Mahwah, NJ: Lawrence Erlbaum Associates; 2013.

45. Hamby SL, Finkelhor D, Ormrod R, Turner H. *The Juvenile Victimization Questionnaire (JVQ): Administration and scoring manual.* Durham, NH: Crimes Against Children Research Center; 2005.

46. Tukey JW. *Exploratory data analysis.* Pearson; 1977.

47. Hubert M, Vandervieren E. An adjusted boxplot for skewed distributions. *Comput Stat Data Anal.* 2008;52(12):5186-5201.

48. Little TD, Slegers DW, Card NA. A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Struct Equ Model.* 2006;13(1):59-72.

49. Little TD, Lopez DF. Regularities in the development of children's causality beliefs about school performance across six sociocultural contexts. *Dev Psychol.* 1997;33(1):165-175.

1. A standard parameterization results when the set of new effects-coded predictors are entered as continuous covariates. Our models had a non-standard parameterization because the new predictors were unnecessarily entered as categorical. This is of no consequence for 0/1 dummy coding, but alters the parameters of the 1st and last level when 1/0/-1 effects codes are present. The model fit and residuals are identical for both parameterizations, however there is a small shift in the intercept parameter in response to two other parameter changes. [↑](#footnote-ref-1)