



# Impact Evaluation of the Alabama Youth Relationship Education Project in Auburn, AL

**Final Impact Evaluation Report for**

Alabama Youth Relationship Education (AYRE)

August 12, 2025

**yPrepared by**

Kim Gregson, PhD, LMFT, Gregson Consulting LLC

Francesca Adler-Baeder, PhD, Auburn University

Julianne McGill, PhD, Auburn University

Menglin Wei, PhD, Auburn University

Ziying Wang, PhD, Auburn University

**Recommended Citation:**

Gregson, K., Adler-Baeder, F., McGill, J., Wei, M., & Wang, Z. (2025). *Impact Evaluation of the Alabama Youth Relationship Education Project in Auburn, AL*. Office of Planning, Research, and Evaluation; Administration for Children and Families.

## **Acknowledgements:**

A heartfelt thank you to the staff, students, facilitators, state partners, and school partners of the Alabama Youth Relationship Education project. Your dedication, enthusiasm, and hard work have been instrumental in the success of this impact evaluation. Most importantly, we deeply appreciate the high school youth who participated in the programs and the study. Your commitment to advancing research on relationship education in the community is truly exemplary, and we value the time you invested in this process.

We are also grateful to Derekh Cornwell and Julius Anastasio from Mathematica for your exceptional technical assistance throughout the project. Additionally, we extend our sincere appreciation to our Federal Program Specialists, Toya Joyner and Meghan Heffron, for your unwavering leadership, dedication, and support at every stage of our project and research.

### Auburn University staff & students (alphabetical):

Benjamin Burke, Melike Coskun, Erin Cooper, Merrill Ann Culverhouse, Jenna Lawlor, Katie Lawrence, Dejah Perrier, Raequon Pinkney, Tonya Rogers, Kaitlynn Slaughter, Je'Kylynn Steen, Brad Thomas, Teresa Wagoner, Paityn Whittington

### Implementation partner sites (alphabetical):

ACES of Elmore County, Circle of Care Center for Families, Family Guidance Center of Alabama, Hope Place Family Resource Center, Sylacauga Alliance for Family Enhancements, The Family Success Center of Etowah County, Tuscaloosa's One Place, Window Seat Creative

### High schools (alphabetical):

Auburn High School, Beauregard High School, Beulah High School, Booker T. Washington High School, Brewbaker Technological Magnet High School, Elmore County High School, Etowah County High School, G. W. Carver High School, Gadsden City High School, Genesis Alternative School, Glencoe High School, Glenwood Academy, Hokes Bluff High School, Holtville High School, Jefferson Davis High School, Lanett High School, Loachapoka High School, Notasulga High School, Opelika High School, Opelika Learning Center, Percy L. Julian (Lee) High School, Reeltown High School, Sardis High School, Sidney Lanier High School, Smith Station High School, Southside High School, Stanhope Elmore High School, Sylacauga High School, Tallassee High School, Valient Cross Academy, Valley High School, West End High School, Wetumpka High School

## **Disclosure:**

The lead author of this report is not affiliated with Auburn University and served as the outside evaluator in collaboration with the co-authors affiliated with Auburn University and the program team.

This publication was prepared under Grant Number 90ZD0002 from the Office of Family Assistance (OFA) within the Administration for Children and Families (ACF), U. S. Department of Health & Human Services (HHS). The views expressed in this report are those of the authors and do not necessarily represent the views or policies of HHS, ACF, or OFA.

## **Structured Abstract: “Impact Evaluation of the Alabama Youth Relationship Education Project in Auburn, AL”**

**Objective.** The Alabama Youth Relationship Education (AYRE) project, a partnership between Auburn University and eight program delivery sites, conducted a longitudinal, quasi-experimental design evaluation assessing the equivalence of effects between two program conditions (i.e., “near-peer” [NP] facilitator model versus “community educator” [CE] facilitator model) delivering relationship education programming to Alabama high school youth.

**Study design.** Over a 2 ½-year period, youth (n = 4,270) enrolled in the impact study and participated in one of two program conditions. Both groups received the same 12-hour school-based AYRE curriculum, but the type of facilitators differed: (1) NP facilitator model, involving college-aged, newly trained facilitators (enrolled youth n = 1,787), and (2) CE facilitator model, involving community agency staff with extensive experience (enrolled youth n = 2,483). Each program condition was implemented at separate, non-overlapping high school sites. Youth completed baseline and follow-up surveys to address primary research questions that hypothesized equivalence between groups at 6 months post-baseline on measures of relationship skills knowledge, dating aggression tolerance, relationship self-efficacy, and self-regulation.

**Results.** The study had low attrition at both individual- and cluster-levels. After creating matched samples via propensity score matching, equivalent effects testing of the matched samples (with covariate-adjusted outcome means) at 6-month follow-up demonstrated that youth in the NP and CE conditions were equivalent on dating aggression tolerance and self-regulation. Youth in the two conditions were not equivalent on relationship skills knowledge or relationship self-efficacy, with NP youth reporting significantly lower relationship skills knowledge and higher relationship self-efficacy, compared with CE youth.

**Conclusion.** Utilizing near-peer educators who are trained and supervised to provide youth relationship education (YRE) is a viable program delivery model, in addition to the previously validated model of providing YRE with trained community educators. While CE facilitators may offer an advantage in delivering relationship knowledge, NP facilitators appear to foster greater confidence among youth in their own relationship abilities. The cost-effectiveness and scalability of the near-peer model, along with its equivalent positive impact on key outcomes and enhanced impact on self-efficacy, suggest strong potential for broad implementation in school-based YRE programs.

# Impact Evaluation of the Alabama Youth Relationship Education Project in Auburn, AL

## I. Introduction

This chapter describes the motivation and goals of the study and gives a high-level overview of the research questions.

### A. Study overview

Unhealthy relationships among youth threaten their physical, social, and emotional wellbeing and can have cascading negative consequences over the trajectory of their own and their families' lives (Banyard & Cross, 2008; Bouchey & Furman, 2003; Mendo-Lazaro et al., 2019; Umberson & Karas Montez, 2010). Unfortunately, Alabama has a persistent history of high levels of family instability, and an overwhelming number of families face challenges to economic self-sufficiency. Because of this, Alabama's youth face tremendous risks to their healthy development and relationships. The potential value for youth relationship education (YRE) is evident, based on state and national data as well as feedback from community interest holders.

Invoking a prevention science approach (Coie et al., 1993), we expected that protective factors, such as evidence-based YRE and links to support services and resources, could be put in place to promote resilience and positive outcomes. Indeed, results of evaluations of YRE document improvements for youth in relationship knowledge, conflict management skills, openness to attending future marriage/relationship education classes, as well as reductions in faulty relationship beliefs and the use of verbal aggression (e.g., Adler-Baeder et al., 2007; Futris et al., 2017; McElwain et al., 2017; Rice et al., 2017; Savasuk-Luxton, et al., 2018). Recent reviews of best practices in YRE (Hawkins, 2018; Simpson et al., 2018) suggest the need for further evaluations of program outcomes, as well as implementation conditions that influence effectiveness. These suggestions align with the learning agenda of the Administration for Children and Families (ACF) and provide the basis for this local evaluation study design. In particular, with established effects of YRE delivered by experienced community educators (i.e., staff from community agencies with substantial experience in training and delivering YRE) and the profound need for YRE as a protective intervention implemented widely to high school youth, we envisioned the utility of delivering YRE via near-peer educators (i.e., newly trained facilitators in the college-age demographic), which may present a cost-efficient option—enhancing the reach to more communities.

While CE facilitators have had years of experience in teaching YRE and research indicates positive outcomes of YRE when participants rate community educators' facilitation knowledge and skills high (Hawkins & Ooms, 2012), we expected the developmental proximity of youth and college-age newly-trained NP facilitators to provide at least an equally positive program experience and outcomes. In past ACF grant cycles, AYRE utilized NP facilitators, and youth provided qualitative feedback that they enjoyed learning from them; however, there are no known prior YRE studies testing whether youth taught by CE and NP facilitators have equivalent outcomes. For the 2020-2025 ACF Healthy Marriage and Responsible Fatherhood (HMRF) grant cycle, the Alabama Youth Relationship Education (AYRE) project conducted an impact evaluation of two program conditions, comparing youth participants who received YRE programming from a "near-peer" (NP) facilitator trained through the University Consortium (i.e., the service of interest) to youth participants in a validated program group who received YRE programming from a community educator (CE) facilitator employed by a family resource center (an evidence-based model). High school students in both conditions of this quasi-experimental design (QED) study received the same YRE content, referred to hereafter as the "AYRE curriculum" (i.e., combination of *Relationship Smarts Plus 4.0*, *Mind Matters*, and *Money Habitudes* curricula; see Chapter II, Section II.B Content for more details about curricula).

## B. Primary research questions

Our primary research questions (RQs) examined whether the impact of CE and NP facilitators delivering the same AYRE curriculum produced equivalent effects on outcomes. Specifically, we assessed whether outcomes were equivalent for participants who received the AYRE curriculum under two program delivery conditions: **(1) NP facilitator model** (i.e., novel program condition; youth participants received the AYRE curriculum from college-age, newly trained NP facilitators, supervised through the University Consortium) and **(2) CE facilitator model** (i.e., validated program condition; youth participants received the AYRE curriculum from CE facilitators, typically with multiple years of experience, employed and supervised by a partner site).

In order to examine sustained outcomes of programming, we assessed whether the NP facilitator model and the CE facilitator model produced equivalent effects on selected outcomes at six months post-baseline (see Chapter IV, Section C for explanation of equivalent effects). Four targeted outcomes, which covered individual and relational domains of functioning, were selected to align with the AYRE curriculum content taught in both program conditions. Specifically, the RQs assessed the equivalent effects of the two facilitator models on relationship skills knowledge, dating aggression tolerance, and relationship self-efficacy (all corresponding with *Relationship Smarts Plus 4.0* curriculum), as well as self-regulation (in line with both *Relationship Smarts Plus 4.0* and *Mind Matters* curricula).

In considering the hypothesized effects, we note that previous evaluation studies of YRE predominantly utilized experienced community educators or teachers and established program impacts for youth in both relational and individual domains (Adler-Baeder et al., 2007; Futris et al., 2017; Hawkins, 2018; McElwain et al., 2017; Rice et al., 2017; Savasuk-Luxton, et al., 2018). Building on the evidence base for YRE, we sought to evaluate whether near-peer facilitators could achieve similar established outcomes to community educators, potentially expanding the versatility of YRE programming via low-cost, broad-reach implementation methods. Experienced community educators have the advantage of years of experience in YRE facilitation and classroom management; however, newly trained near-peer facilitators have the advantage of developmental proximity to youth and may be viewed as more relatable, likely providing an equally positive learning environment. Previous research on facilitator congruence in adult HMRE found some benefit for having a “like other” facilitator (Bradford et al., 2012). Thus, we expected the impact of the two program conditions on specified outcomes to be equivalent. The primary RQs are as follows:

1. Is the impact of the AYRE program on *relationship skills knowledge* equivalent at six months post-baseline between youth who received the program in the NP facilitator model and youth who received the program in the CE facilitator model?
2. Is the impact of the AYRE program on *dating aggression tolerance* equivalent at six months post-baseline between youth who received the program in the NP facilitator model and youth who received the program in the CE facilitator model?
3. Is the impact of the AYRE program on *relationship self-efficacy* equivalent at six months post-baseline between youth who received the program in the NP facilitator model and youth who received the program in the CE facilitator model?
4. Is the impact of the AYRE program on *self-regulation* equivalent at six months post-baseline between youth who received the program in the NP facilitator model and youth who received the program in the CE facilitator model?

## II. Program conditions

This chapter describes the focal population, program descriptions for the NP facilitator model (i.e., novel program condition) and the CE facilitator model (i.e., validated program condition). Additionally, it describes the implementation RQs about the delivery of NP and CE facilitator model conditions.

## A. Focal population

The AYRE project's target population for both NP and CE conditions was youth, grades 9-12, in high schools across Alabama. (For details about sample characteristics, see Section IV.B of this report.)

## B. Description of NP facilitator model as intended

Youth participants in the NP facilitator model received the AYRE curriculum from trained NP facilitators, who were overseen and trained by staff at four program sites across Alabama. Program sites which oversaw NP and CE facilitators did not overlap, and youth were participants in each program condition based on high school location (served by local program site). See Table II.1 for detailed description of the NP facilitator model condition.

**Components.** In the NP facilitator model, youth received the 12-hour workshop series facilitated by two trained NP educators, delivered in high school classes (e.g., Health & Wellness, Career & Business Education, Family & Consumer Science) in a group setting with other youth. NP facilitators taught the AYRE curriculum to youth during regular school hours in the high schools that partnered with the NP facilitator model program sites. Facilitators provided a list of local and national resources to all participants during the workshop series.

**Content.** In both NP and CE facilitator models, youth received the same evidence-informed, evidence-based AYRE curriculum content. The AYRE curriculum, intentionally selected from the *Relationship Smarts PLUS 4.0* (Pearson, 2018), *Mind Matters* (Curtis & Stolzenbach, 2017), and *Money Habitudes for Teens* curricula (Solomon, 2009) focuses on the following healthy relationship skills: communication, conflict resolution and problem-solving, stress and anger management, financial decision-making, self-regulation and mindfulness skills, and the effects of distress on relationship functioning. The scripted AYRE curriculum was delivered using brief informational sessions, individual and group activities, popular media videos, and discussion, all aimed at engaging, informing, and emphasizing the benefits of healthy relationships. Facilitators provided relevant handouts so students could keep records and take materials home.

**Dosage and implementation schedule.** For both NP and CE facilitator models, this 12-hour workshop series involved 9-14 sessions occurring over a minimum of 14 days between the first and last workshop session. All program sites coordinated with participating high schools, and thus there was some variability in the number, duration, and frequency of sessions due to variations in both high school schedules and partnering teachers' classroom schedules. To address the variations in high school schedules (e.g., 1.5-hour block high school schedule, 50-minute traditional high school schedule, etc.), several "AYRE program delivery plans" were developed during the startup period of this grant cycle. For each program delivery plan, the full AYRE curriculum content was delivered over 12 hours of program delivery, but adjustments were made for varying high school class lengths; importantly, the class series did not vary in program content or total hours of program delivery. To address the variations in partnering teachers' classroom schedules, program sites coordinated with partnering teachers at the beginning of each semester to determine the schedule of delivery (e.g., one class session per week, daily class sessions on subsequent school days, etc.). Overall, about two-thirds of the workshop series lasted 10-15 weeks (with approximately 1 class session per week) and one-third of the workshop series lasted about 3 weeks (with daily class sessions on subsequent school days).

**Delivery mode.** In the NP facilitator model, each workshop series was delivered at local Alabama high schools associated with these four program sites: Auburn University (AU; Auburn, AL), Window Seat Collective (Montgomery, AL), Circle of Care Center for Families (Valley, AL) and Family Guidance Center (Montgomery, AL). Trained NP facilitators at each site delivered 12 hours of programming in high schools.

**Staff characteristics, education, and training.** The NP program sites (listed above under Delivery mode) oversaw a university consortium of NP facilitators from nearby universities who

implemented the primary workshop at local partnering high schools. Most NP facilitators were undergraduate students at a local university (or around the age of undergraduate students).

At the beginning of each high school semester (i.e., fall, spring), each NP facilitator participated in a comprehensive initial one-day training, as well as 2-3 hours each week of training classes (for 4-5 weeks) on AYRE curriculum and the evaluation study prior to implementation. At that point, NP facilitators began implementing the program in high school classes, and they continued to participate in ongoing weekly group training classes (led by program staff). Overall, these trainings focused on practicing facilitating the AYRE curriculum through “teach backs” (i.e., facilitation practice session of each lesson) and learning about topics such as group engagement skills, classroom management, key facilitation skills, and implementing an evaluation study.

AU AYRE staff provided weekly monitoring and regular assistance to all NP facilitators and program staff. In addition, the local evaluation team interacted frequently with NP facilitators and program staff: collecting monthly reporting, conducting biannual evaluation trainings, creating thorough written reports of enrollment, program/survey completion and sample characteristics, and offering routine technical assistance via virtual meetings. Open communication lines (i.e., emails, phone calls, Box file sharing) were used on a weekly basis. The local evaluation team and program staff regularly monitored program delivery fidelity and quality via routine site visits and online fidelity checklists, which were completed by NP facilitators after each of the class sessions in every workshop series. Additionally, the local evaluation team and program staff provided ongoing program development opportunities via scheduled virtual meetings about program implementation, evaluation design, data collection, participant engagement, and evaluation feedback.

See Table II.2 for details about staff characteristics in each condition.

**Demographic comparison between NP and CE facilitators.** To test demographic differences between NP facilitators (n=82) and CE facilitators (n=31) who facilitated workshops during the impact study period, we used *t*-tests and Hedge’s *g* effect sizes for continuous variables and chi-square tests and Cox’s index effect sizes for binary variables (see Appendix B, Table B.5). Compared with CE facilitators, the NP facilitators were more proportionally female (NP 82.9% female; CE 71.0% female; Cox’s index effect size = .42), younger (NP  $M_{age}=22.41$ ,  $SD=5.22$ ; CE  $M_{age}=43.30$ ,  $SD=14.76$ ; Hedge’s *g* effect size = 2.36), lower educated (NP 78.0% without a 4-year degree; CE 19.4% without a 4-year degree; Cox’s index effect size = 1.65), not married (NP 92.7% unmarried; CE 25.8% unmarried; Cox’s index effect size = 2.20), and not parents (NP 91.5% not parents; CE 25.8% not parents; Cox’s index effect size = 2.04). NP and CE facilitator groups were similar on race (NP 64.6% white; CE 64.5% white; Cox’s index effect size = .00).

### C. Description of the CE facilitator model as intended

Youth participants in the CE facilitator model received the AYRE curriculum from trained CE facilitators (i.e., regional extension agents or family life educators), at five other program sites across Alabama. Program sites which oversaw NP and CE facilitators did not overlap, and youth were participants in each program condition based on high school location (served by local program site). See Table II.1 for detailed description of the CE facilitator model condition.

**Components.** In the CE facilitator model, youth received the same 12-hour workshop series facilitated by two trained CE educators, delivered in high school classes (e.g., Health & Wellness, Business & Career Planning, Family & Consumer Science) in a group setting with other youth. CE facilitators taught the AYRE curriculum to youth during regular school hours in the high schools that partnered with the CE facilitator model program sites. Facilitators provided a list of local and national resources to all participants during the workshop series.

**Content.** See above Section II.B Content



**Dosage and implementation schedule.** *See above Section II.B Dosage and implementation schedule*

**Delivery mode.** In the CE facilitator model, each workshop series was delivered at local Alabama high schools associated with these five program sites: ACES of Elmore County (Wetumpka, AL), Family Success Center of Etowah County (Gadsden, AL), Hope Place Family Resource Center (Brewton, AL), Sylacauga Alliance for Family Enhancement (Sylacauga, AL), and Tuscaloosa's One Place (Tuscaloosa, AL). Two trained CE facilitators at each site delivered the 12 hours of programming.

**Staff characteristics, education, and training.** The CE program sites (listed above under Delivery mode) oversaw CE facilitators who implemented the primary workshop at local partnering high schools. Most CE facilitators had professional backgrounds working in family services.

Each CE facilitator was given a comprehensive 2-day training on AYRE curriculum and the evaluation study prior to implementation. CE facilitators also participated in an annual day of refresher training in the AYRE curriculum and evaluation study methods, as well as monthly ongoing trainings on topics such as group engagement skills, classroom management, current relationship education research, and implementing an evaluation study.

Importantly, most CE facilitators and program staff had prior experience enrolling community participants in evaluation studies. They understood the purpose of a rigorous design, necessity of curriculum fidelity, and need for data collection integrity. AU AYRE staff provided weekly monitoring and regular assistance to all CE facilitators and program staff. In addition, the local evaluation team interacted frequently with CE facilitators and program staff collecting monthly reporting, conducting biannual evaluation trainings, creating thorough written reports of enrollment, program/survey completion and sample characteristics, and offering routine technical assistance via virtual meetings. Open communication lines (i.e., emails, phone calls, Box file sharing) were utilized on a weekly basis. The local evaluation team and program staff regularly monitored program delivery fidelity and quality via routine site visits and online fidelity checklists, which were completed by CE facilitators after each of the class sessions in every workshop series. Additionally, the local evaluation team and program staff provided ongoing program development opportunities via scheduled virtual meetings about program implementation, evaluation design, data collection, participant engagement, and evaluation feedback.

See Table II.2 for details about staff characteristics in each condition.

**Table II.1. Description of intended NP and CE facilitator model conditions, including components, curriculum and content, dosage and schedule, delivery, and focal population**

| Component                     | Curriculum and content                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Dosage and schedule                                                                                                                                                                                                                                                         | Delivery                                                                                                                                                                           | Focal population                                   |
|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|
| NP facilitator model          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                             |                                                                                                                                                                                    |                                                    |
| Relationship skills workshops | <p>-AYRE curriculum (selected from <i>Relationship Smarts PLUS 4.0</i>, <i>Mind Matters</i>, and <i>Money Habitudes for Teens</i> curricula) covering the following healthy relationship topics: communication skills, conflict resolution and problem-solving, stress and anger management, financial decision-making, self-regulation and mindfulness skills, and the effects of distress on relationship functioning</p> <p>-A list of local and national resources was provided to all participants</p> | <p>-12-hour workshop, with 9-14 sessions occurring over a minimum of 14 days between the first and last workshop session</p> <p>-Some variability in number, duration, and frequency of sessions, according to the schedule of the partnering high school &amp; teacher</p> | <p>-Group lessons in classrooms at local high schools partnered with 4 different AYRE NP partner sites across Alabama</p> <p>-Two NP facilitators provided the workshop series</p> | Youth (grades 9-12) in high schools across Alabama |
| CE facilitator model          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                             |                                                                                                                                                                                    |                                                    |
| Relationship skills workshops | <p>-AYRE curriculum (selected from <i>Relationship Smarts PLUS 4.0</i>, <i>Mind Matters</i>, and <i>Money Habitudes for Teens</i> curricula) covering the following healthy relationship topics: communication skills, conflict resolution and problem-solving, stress and anger management, financial decision-making, self-regulation and mindfulness skills, and the effects of distress on relationship functioning</p> <p>-A list of local and national resources was provided to all participants</p> | <p>-12-hour workshop, with 9-14 sessions occurring over a minimum of 14 days between the first and last workshop session</p> <p>-Some variability in number, duration, and frequency of sessions, according to the schedule of the partnering high school &amp; teacher</p> | <p>-Group lessons in classrooms at local high schools partnered with 5 different AYRE CE partner sites across Alabama</p> <p>-Two CE facilitators provided the workshop series</p> | Youth (grades 9-12) in high schools across Alabama |

**Table II.2. Staff characteristics, education, initial training, and ongoing training to support NP and CE facilitator model components**

| Component                     | Staff characteristics, education, and initial training                                                                                                                                                                                                                                                                                                                                                                                                                                         | Ongoing staff training                                                                                                                                                                                                                                                                                                       |
|-------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| NP facilitator model          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                              |
| Relationship skills workshops | NP facilitators were proportionally more female, younger, less educated, fewer married, and fewer indicating that they were parents than CE facilitators. Most of them were currently in the process of completing an undergraduate degree. Prior to implementation, they received one day of initial training, as well as 2-3 hours each week of training meetings (for 4-5 weeks) on the AYRE curriculum and the evaluation study in addition to continuous updates on relevant information. | NP facilitators participated in weekly group classes (led by program staff) to practice facilitating AYRE curriculum and to learn about topics such as group engagement skills, classroom management, relationship education research, and implementing an evaluation study.                                                 |
| CE facilitator model          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                              |
| Relationship skills workshops | CE facilitators were proportionally more mixed sex (male/female), higher educated, more married, and more indicating that they were parents than NP facilitators. Most of them had professional backgrounds working in family services. Prior to implementation, they received two days of initial training in the AYRE curriculum and the evaluation study in addition to continuous updates on relevant information.                                                                         | CE facilitators received a day of annual refresher training in the AYRE curriculum and evaluation study methods from AU program and local evaluation staff. They also received monthly trainings about group engagement skills, classroom management, relationship education research, and implementing an evaluation study. |

## D. Implementation research questions about the program conditions

This section presents the research questions (RQs) to understand the implementation of the AYRE project. Per ACF HMRF guidelines, we assessed implementation elements of fidelity, dosage, quality, engagement, and context in the two program conditions: NP facilitator model and CE facilitator model. Implementation RQs, grouped by implementation elements, are listed below:

1. **Fidelity:** For each condition, according to facilitator reports after each class session, what average percentage of curriculum content was taught?
2. **Dosage:** For each condition, what percentage of individuals attended: 0%, 50%, 90%, or 100% of the 12-hour workshop series?
3. **Dosage:** For each condition, what is the average number of programming hours received?
4. **Quality:** For each condition, what is the average participant post-program rating of facilitation quality?
5. **Engagement:** For each condition, what is the average participant post-program rating of self-engagement during the class series?
6. **Context:** For each condition, what type of high school classes were the AYRE workshop series delivered (by percentage)?

## III. Study design

This chapter describes the sample formation, research design, and data collection for both the impact and implementation analyses. It also describes the methods used to analyze the data as part of the study. For more detailed descriptions of each of the sections below, see Appendix B.

### A. Evaluation enrollment and assignment to study conditions

**Recruitment.** Both NP and CE program sites recruited participants from high schools through relationships with teachers in subjects like Health & Wellness, Business & Career Planning, and Family & Consumer Science. Participants were enrolled through parent consent and youth assent.

**Cohort Structure & Demographics.** The study used 20 separate cohorts, identified by the month and year classes began, and aimed to balance youth demographics, such as sex and race, across NP and CE groups.

**Cluster Assignment & Balance.** Because the impact study used a QED approach with cluster assignment at the high school level, efforts were made at all phases of planning and implementation to ensure demographic balance between the NP and CE groups by paying attention to county settings (i.e., urban/rural) and student demographics, including a target of 40% non-white students for both groups.

**Monitoring & Adjustments.** The evaluation team closely monitored participant demographics and made adjustments to high school partnerships and enrollment numbers during the first year to achieve comparability across groups, and the team continued to monitor these variables quarterly throughout the study.

**Community Partnerships.** The AYRE program's long-standing partnerships with local high schools and community organizations played a key role in the success of the study. Communication and coordination between program staff and school personnel were ongoing throughout the implementation phase.

**Enrollment targets.** The targeted AYRE impact study enrollment was 4,200 youth, given AYRE's history of outreach and program delivery over a similar time period. Initially, we expected a 50/50 balance of youth in the NP and CE facilitator conditions; however, over the course of implementation, we recognized that it was more realistic to expect that the more experienced CE program sites would serve 60% of the target sample and the newer NP program sites would serve 40% of the target sample. Thus, the target enrollment was 1,680 youth in the NP facilitator model and 2,520 youth in the CE facilitator model. By April 2024 (i.e., 6 months earlier than the planned end date), AYRE had successfully enrolled more than the targeted sample for the impact study (i.e., targeted sample  $n = 4,200$  youth; enrolled sample  $n = 4,270$  youth), and ACF gave approval to stop enrolling new youth into the impact study. See Appendix B, CONSORT diagram Figures B.1 and B.2 for details about the analytic samples. See Appendix E, Section E.2 for power analysis of final analytic samples for each RQ.

**Participant eligibility criteria.** Participant eligibility criteria were the same for both NP and CE facilitator models. Participants were eligible to participate if they were high school youth (grades 9-12) enrolled in participating classes in a participating Alabama high school.

**Special recruitment and enrollment procedures.** There were no additional criteria for recruiting and selecting the sample beyond the eligibility criteria.

**Consent process.** The AU Institutional Review Board (IRB) approved all study procedures originally on 3/23/21 with no expiration date (Protocol # 21-007 EP 2103), and two subsequent revisions to the original protocol were approved on 1/3/23 (i.e., additional contact info requested on informed consent & addition of COVID permission letters from schools) and 7/7/23 (i.e., change in project location for storing data & added language about archiving data on the consent), with no expiration dates.

The consent process was identical across conditions and involved informing students about the evaluation, confidentiality, and voluntary participation using a scripted explanation, followed by distribution of IRB-approved consent/assent letters via hard copy or secure electronic link. Only students with signed consent/assent forms were included in evaluation activities, and their information was securely stored in Ripple Science, a secure online participant management system, with restricted access by trained staff.

**Quasi-experimental design and research group formation.** This cluster QED impact evaluation used two naturally occurring groups of participants. Groups were formed based on student enrollment in classes (e.g., Health & Wellness, Career & Business Education, Family & Consumer Science) at local partnering high schools. Any student enrolled in the participating high school class was eligible to participate in the AYRE impact study. High schools were grouped into program conditions (NP facilitator model or CE facilitator model) by the program site serving that school (see above Chapter II, Section B & C for details about program sites for each condition). Thus, consent and baseline data collection occurred after designation of condition. Importantly, all data analyses used propensity score matching to render the two groups similar so that analytic comparisons could be made (see Chapter IV, Section IV.C. below for details on analytic procedures).

## **B. Outcome measures**

Outcome measures for all primary RQs were collected on AYRE local evaluation surveys (via Qualtrics) and nFORM (Information, Family Outcomes, Reporting, and Management) surveys; nFORM is a secure, federally managed participant management and data collection system used by HMRF-funded grantees to track enrollment, service delivery, and outcome measures for program evaluation and performance reporting. Outcome measures for primary RQs include:

**RQ 1: Relationship skills knowledge.** Relationship skills knowledge was measured by the 10-item multiple choice *Relationship Smarts Plus 4.0* quiz (adapted from the Dibble Institute), assessing knowledge about healthy relationship skills taught in the *Relationship Smarts Plus 4.0* curriculum. Prompt and example item: "The following are questions about personal relationships that you'll be learning about

in the Relationship Smarts classes. (1) Your day-to-day decisions, the negative and positive influences in your life, and your relationships will shape your: (a) Possible selves, (b) Multiple personalities, (c) Maturity, or (d) Romantic attractions.” For each of the 10 multiple choice items, 1 answer is correct and 3 answers are incorrect. Each item was recoded from the original 4 response options into dichotomous variables (i.e., 0 = incorrect answer, 1=correct answer). Sum composite scores of the number of total correct items were calculated, and higher scores indicate better knowledge about healthy relationship skills.

**RQ 2: Dating aggression tolerance.** Dating aggression tolerance was measured by a 4-item scale selected for nFORM surveys, assessing participants’ acceptance of aggressive behaviors toward romantic partners. Prompt and example item: “How much do you agree or disagree with the following statements about relationships? A person who makes their partner angry on purpose deserves to be hit.” Values range from 1=“strongly disagree” to 4=“strongly agree”. Mean composite scores of all 4 items were calculated, and higher scores indicate greater acceptance of dating aggression.

**RQ 3: Relationship self-efficacy.** Relationship self-efficacy was measured by a 6-item scale selected for nFORM surveys, assessing participants’ confidence in their ability to use healthy relationship skills. Prompt and example item: “How much do you agree or disagree with the following statements: I have the skills needed for a lasting, stable, long-term relationship.” Values range from 1=“strongly disagree” to 4=“strongly agree”. At baseline, the nFORM entrance survey displayed this 6-item scale to youth who were NOT in a dating relationship. In order to collect data on the full sample, the T1 impact survey displayed this scale to youth who WERE in a dating relationship (to fill in the gap). After baseline data collection was complete, relationship self-efficacy items for non-dating youth from the nFORM entrance survey and for dating youth from the T1 impact survey data were downloaded and merged—such that the full sample (of dating and non-dating youth) had baseline data for this measure. At 6-months post-baseline, the T3 impact survey collected this scale from all youth (dating and non-dating). Mean composite scores of all 6 items were calculated, and higher scores indicate greater relationship self-efficacy.

**RQ 4: Self-regulation.** Self-regulation was assessed as a mean composite of 10 items: 4 items from the Adolescent Self-Regulatory Inventory (ASRI; Moilanen, 2007) and 6 items from the Social-Emotional Health Survey-Secondary (SEHSS; Furlong et al., 2018). Both measures assess youth self-regulation in emotional, behavioral, and cognitive domains. Prompt and example item from the ASRI: “For each statement below, please select a response that best describes you: I can calm myself down when I’m excited or all wound up.” Prompt and example item from the SEHSS: “For each statement please select a response that describes you: “I can wait for what I want.” For both measures, values range from 1=“not at all true for me” to 5=“really true for me”. Mean composite scores of all 10 items were calculated, and higher scores indicate better self-regulation.

See Tables III.1a and III.1b for a description of outcome measures (i.e., citation, items, values, internal consistency), source, and timing of measure collection. See Appendix B, Table B.3 for all outcome measure items and Appendix D, Table D.1 for internal consistency of outcome measures across time points.

### **C. Implementation measures**

See Table III.2 for a description of all implementation measures, and Appendix B, Table B.4 for all implementation measure items.

**Table III.1a. Outcome measures used to answer primary research questions 1 & 2 of the impact analysis**

| Research question # | Outcome name                  | Description of the outcome measure and its properties                                                                                                                                                                                                                                  | Source of the measure                                                                                                                                                                                                                                                                                                  | Timing of measure                   |
|---------------------|-------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|
| 1                   | Relationship skills knowledge | Relationship skills knowledge was assessed using the sum of 10 multiple choice items measuring youth's understanding of healthy relationship skills taught in the <i>Relationship Smarts Plus 4.0</i> curriculum, with higher scores indicating better knowledge (T1 M=6.03, SD=1.62). | Baseline: T1 impact survey (Qualtrics)<br><br>6-Months Post-Baseline: T3 impact survey (Qualtrics)                                                                                                                                                                                                                     | Baseline ) & 6-months post-baseline |
| 2                   | Dating aggression tolerance   | Dating aggression tolerance was assessed using a mean of 4 Likert scale items measuring youth's acceptance of aggressive behaviors toward romantic partners, with higher scores indicating greater acceptance of dating aggression (range across time points $\alpha = .72 - .83$ ).   | Baseline: nFORM entrance survey (nFORM online platform)<br><br>6-Months Post-Baseline: T3 impact survey (Qualtrics)                                                                                                                                                                                                    | Baseline & 6-months post-baseline   |
| 3                   | Relationship self-efficacy    | Relationship self-efficacy was assessed using a mean of 6 Likert scale items measuring youth's confidence in their ability to use healthy relationship skills, with higher scores indicating greater relationship self-efficacy (range across time points $\alpha = .78 - .82$ ).      | Baseline: nFORM entrance survey (nFORM online platform) measured relationship self-efficacy among non-dating youth; T1 impact survey (Qualtrics) measured relationship self-efficacy among dating youth<br><br>6-Months Post-Baseline: T3 impact survey (Qualtrics) measured this on all youth (dating and non-dating) | Baseline & 6 months post-baseline   |
| 4                   | Self-regulation               | Self-regulation was assessed using a mean of 10 Likert scale items measuring youth self-regulation in emotional, behavioral, and cognitive domains, with higher scores indicating better self-regulation (range across time points $\alpha = .74 - .81$ ).                             | Baseline: T1 impact survey (Qualtrics)<br><br>6-Months Post-Baseline: T3 impact survey (Qualtrics)                                                                                                                                                                                                                     | Baseline & 6-months post-baseline   |

**Table III.2. Measures used to address implementation research questions**

| Implementation element | Research question                                                                                                                                                                                        | Measures                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fidelity               | 1. For each condition, according to facilitator reports after each class session, what average percentage of curriculum content was taught?                                                              | For each condition, average percentages were calculated as the average number of curriculum activities checked by the lead facilitator on the fidelity checklist (after each class session of each workshop series) out of the total number of planned curriculum activities. Percentages indicate how much of the intended program was actually delivered.                                                                                                                                                                       |
| Dosage                 | 2. For each condition, what percentage of individuals attended: 0%, 50%, 90% or 100% of the 12-hour workshop series?<br>3. For each condition, what is the average number of programming hours received? | For each condition, we calculated frequencies (percentage) of individuals who did not attend any (0%) sessions, attended 50%, 90%, or 100% of the 12-hour workshop series (from nFORM session attendance information). We also calculated mean number of programming hours received by condition.                                                                                                                                                                                                                                 |
| Quality                | 4. For each condition, what is the average participant post-program rating of facilitation quality?                                                                                                      | For each condition, mean composites were calculated for participant-reported facilitation quality, a 5-item subscale of the Youth Program Classroom Education Environment Scale (YPCEES; Adler-Baeder et al., 2025) on AYRE immediate post-program impact surveys. Higher scores indicate better facilitation quality (e.g., explaining material, managing time, using appropriate examples). This subscale has excellent reliability in the current sample ( $\alpha = .93$ ).                                                   |
| Engagement             | 5. For each condition, what is the average participant post-program rating of self-engagement during the class series?                                                                                   | For each condition, mean composites were calculated for participant-reported individual engagement, a 3-item subscale of the YPCEES (Adler-Baeder et al., 2025) on AYRE immediate post-program impact surveys. Higher scores indicate greater self-engagement in the class series (e.g., sharing in group discussions, feeling connected to other participants). This subscale has good reliability in the current sample ( $\alpha = .83$ ).                                                                                     |
| Context                | 6. For each condition, in what type of high school classes were the AYRE workshop series delivered (by percentage)?                                                                                      | For each condition, we calculated percentages of types of high school classes in which AYRE workshop series were delivered from the staff-reported class type variable. Response options included: Career & Business Education (i.e., Business Essentials, Career Prep, Future Farmers of America), Family & Consumer Sciences, Health & Wellness (i.e., Health, Physical Education), Humanities & Social Studies (i.e., History, Social Studies, Bible), or Other (i.e., Afterschool, Alternative School, Freshman Orientation). |



## D. Data collection

This section details the data collection procedures for the impact analysis (both NP and CE facilitator model conditions), including engagement strategies used to retain participants. Next, we describe data collection procedures for the implementation analysis for both conditions.

### 1. Impact data collection

All data collection procedures were approved in the AU IRB study protocol (see above Section III.A). Data collection procedures for youth in both program conditions were identical. After youth and their parents/guardians assented and consented, respectively, to participate in the AYRE impact study, program staff held a “baseline data collection day” in the high school classroom, prior to the 12-hour workshop series. Consented participants completed three surveys at baseline: nFORM Applicant Characteristics Form, nFORM Entrance survey, and AYRE baseline impact survey (T1). After completing the 12-hour workshop series, consented participants completed two surveys on an “immediate post-program data collection day”: nFORM Exit survey and AYRE immediate post-program impact survey (T2); both the nFORM Exit and T2 impact surveys were not used in this final evaluation report. Six months after baseline data collection, participants were emailed and texted a link to the AYRE 6-month impact survey (T3). Baseline and immediate post-program survey collection occurred in high school classrooms and was overseen by partner staff, who were thoroughly trained by the evaluation team to comply with IRB-approved data collection procedures. Six-month follow-up survey data collection was overseen by the evaluation team and occurred on participants’ own devices. To decrease attrition, compensation in the form of a \$25 gift card was offered to participants who completed the final T3 survey.

The AYRE impact surveys (T1 and T3) and nFORM entrance surveys collected measures of the outcomes of interest for primary RQs 1-4 for participants in both program conditions. See Appendix B, Table B.1 for key features of data collection for the impact analysis and Appendix B, Section B.3 for detailed description of impact data collection.

**Tracking participants and reducing attrition.** Low follow-up response rates are common among community-based studies, particularly with high school youth. To promote engagement and optimize survey response rates, we obtained a variety of modes of contact (i.e., phone, email, parent phone, parent email) from youth and parents on the informed consent/assent form. For the T3 impact survey, the local evaluation team sent “save the date” notices (via email and text) one week prior to sending survey links, reminding participants about the upcoming survey and asking about any changes to contact information. After the local evaluation team sent the T3 impact survey link, they sent multiple reminder texts/emails/phone calls during the 4-week “open” window for survey completion. Additionally, during each “open” window for T3 impact survey completion, the evaluation team sent encrypted lists of participants who had not yet completed the survey to program staff, who directly contacted the relevant participants to remind them to complete before the survey window closes. In an added effort to engage and retain participants through the final T3 impact survey, program staff advertised the AYRE social media accounts for youth to follow while they were in the 12-hour class series. Program staff frequently posted on AYRE social media accounts information about healthy relationship skills, relevant local resources, and reminders about T3 impact survey timing. Following best practices for survey research and to enhance response rates, youth participants were compensated for their participation in the local impact evaluation after completing the T3 impact survey via \$25 electronic gift cards delivered to their email.

To monitor survey completion, the evaluation team used the Ripple software program to organize and track participants by program condition. Each month throughout the duration of the impact study, we calculated overall attrition of the entire sample (both NP and CE facilitator models), as well as differential attrition between the two conditions. The evaluation team and program team focused ongoing Continuous Quality Improvement efforts on minimizing overall and differential attrition in accordance with the What Works Clearinghouse (WWC) Standards (2022). See Appendix C, Table C.3 for overall and differential

attrition rates at the individual- and cluster-levels. See Appendix B for CONSORT diagram Figures B.1 and B.2 tracking the sample (separately by condition) at the individual and cluster-levels.

## 2. Implementation data collection

A variety of data sources were used to address the implementation RQs (see Appendix B, Table B.2 for an overview of data sources, timing of data collection, and responsible parties for implementation measures). To address **Fidelity RQ #1**, after each class session in all workshop series (i.e., 156 NP facilitator model workshop series, 165 CE facilitator model workshop series), the lead facilitator completed a paper fidelity checklist indicating which of the required curriculum elements were taught during that class session. At the end of each class series, paper fidelity checklists were entered by evaluation staff into an SPSS database. For **Dosage RQs #2 and #3**, AU AYRE staff and partner site staff tracked workshop attendance in nFORM for youth at all sessions during each of the 12-hour workshop series for both NP and CE facilitator models. Youth participants in both conditions completed T2 impact surveys at immediate post-program, including two subscales of the Youth Program Classroom Education Environment Scale (YPCEES; Adler-Baeder et al., 2025): the 5-item facilitation quality subscale (used for **Quality RQ #4**) and 3-item individual engagement subscale (used for **Engagement RQ #5**). Finally for **Context RQ #6**, AU AYRE staff and partner site staff recorded the type of high school class (e.g., Career & Business Education, Family & Consumer Sciences, Health & Wellness, etc.) in which each AYRE workshop series was delivered.

## IV. Analytic methods

This chapter describes the analytic sample, measures, baseline equivalence, and sample characteristics. It also describes the analytic methods for estimating program impacts on the primary research questions.

### A. Analytic sample

This section outlines the construction of the analytic sample, including crossover, attrition, missing data, and determining the analytic samples for each primary RQ. For details about data preparation and cleaning, reference Appendix D.

**Condition Crossover.** For impact studies it is critical to prevent crossover between conditions and re-enrollment into the study by a prior participant. The local evaluation team and program staff monitored participant enrollment and class attendance via nFORM and frequent communication with partner staff. There was almost no likelihood for enrolled participants to switch program conditions, which would have required the student to move high schools during the impact study period (since conditions were assigned at the school level). However, to be cautious, the data manager and evaluation team closely monitored nFORM enrollment to make sure that no participants re-enrolled and that any possible duplicate cases were discovered before enrollment into the impact study. If participants attempted to re-enroll, they were informed that they were not eligible to participate in the impact study again (until the impact study concluded in Fall 2024). Throughout the duration of the impact study, there was no condition crossover; that is, no participants in the NP facilitator model participated in the CE facilitator model, and vice versa.

**Survey Response Rates.** Note that for this study using researcher-collected data (rather than administrative outcome data available for each program participant without consent), youth were enrolled in the study when they consented to participate in the evaluation. The consented, per the guiding IRB guidelines for voluntary research participation (approved AU IRB Protocol # 21-007 EP 2103), are the analytic sample and the only ones from whom it is possible to collect data at each time point. Therefore, we calculated survey response rates as survey respondents out of number consented and attrition rates as those who did not complete a survey out of number consented to participate in the evaluation study. See Table IV.1 for cluster and individual sample sizes and response rates by condition and primary outcome.

**Attrition.** We employed an Intent-to-Treat (ITT) design, recognized as a highly rigorous evaluation method that provides the most reliable efficacy outcomes when assessing a program in a real-world context (Weiss & Jacobs, 2008; Wood et al., 2014). We calculated group comparisons based on assignment to either of the two program delivery conditions (i.e., NP facilitator model or CE facilitator model) rather than actual participation. Prior to defining analytic samples and addressing missing data through imputation, we monitored both overall and differential attrition at the cluster and individual levels, in accordance with WWC Standards (2022) optimistic boundary for differential attrition between conditions (per ACF guidelines that the WWC optimistic boundary can be used for youth evaluations conducted during the regular school day in high school classrooms). All individual- and cluster-level T1-T3 differential attrition rates between program conditions were below the respective WWC (2022) optimistic boundaries at the individual-level (i.e., overall attrition = 36%; corresponding WWC optimistic boundary for a study with 36% overall attrition = 5.3%; differential rates between conditions for T1=1.8%; T2=1.4%; T3=3.2%) and cluster-level (i.e., overall attrition = 0%; WWC optimistic boundary for a study with 0% overall attrition = 10.0%; differential rates between conditions for T1=0.0%; T2=0.0%; T3=0.0%). Thus, this study is classified as a low attrition study at both individual- and cluster-levels, according to WWC Standards (2022). See Appendix C, Table C.3 for overall and differential attrition rates at both individual- and cluster-levels, as well as details about WWC Standards (2022).

**Missing data.** We analyzed missing data patterns at both the survey and item levels. First, for the overall analytic sample, cases without any T1 survey data were removed through listwise deletion. For each primary RQ analytic sample, we assessed item-level missingness for each scale measuring an outcome. Given that all RQs relied on composite scale outcome measures, cases with 20% or fewer missing items on a composite scale were retained, and missing scale items were imputed (see more in this section below). Cases with more than 20% missing items on a composite scale were removed via listwise deletion. This process was conducted separately for each primary RQ and its corresponding composite scale outcome measure. To ensure the robustness of results with imputed data, sensitivity analyses were performed using complete-case analytic samples. Detailed missingness rates for each outcome by condition can be found in Appendix D, Table D.4.

For those cases missing 20% or fewer of the items on a composite scale, we conducted multiple imputation using the “mice” (i.e., Multivariate Imputation by Chained Equations) package in R (Van Buuren & Groothuis-Oudshoorn, 2011; R Core Team, 2021), which is a classification and regression trees (CART) method of using observed values in the dataset as covariates to predict multiple values for each missing value. The resulting dataset included 10 imputed datasets of composite scale scores, and we pooled the mean at the item level of the 10 imputed datasets into a single dataset for data analyses (see Appendix D for details on analytic process of imputing missing data).

**Determining analytic samples for each primary RQ.** See Table IV.1 for cluster and individual sample sizes by condition and primary outcome; see Appendix B for CONSORT diagram Figures B.1 and B.2. Our analytic samples were determined separately for each primary RQ, due to survey- and item-level missingness for relevant surveys and outcomes. Primary RQ analytic sample descriptions by RQ# below:

1. The analytic sample included individual participants in both conditions who had 100% items on the Relationship Knowledge outcome (since this measure was an index composite of correct answers to 10 quiz items) on the T1 impact survey and the T3 impact survey (see Appendix B, CONSORT diagram Figure B.1).
2. The analytic sample included individual participants in both conditions who had 100% items on the Dating Aggression Tolerance outcome (since this scale included only 4 items) on the nFORM Entrance survey and the T3 impact survey (see Appendix B, CONSORT diagram Figure B.1).
3. The analytic sample included individual participants in both conditions who had  $\geq 80\%$  items on the Relationship Self Efficacy outcome on the nFORM Entrance survey (for non-dating youth) or T1 impact survey (for dating youth) and the T3 impact survey (for both dating and non-dating youth; see Appendix B, CONSORT diagram Figure B.1).
4. The analytic sample included individual participants in both conditions who had  $\geq 80\%$  items on the Self-Regulation outcome on the T1 impact survey and the T3 impact survey (see Appendix B, CONSORT diagram Figure B.1).

**Table IV.1. Cluster and individual sample sizes by condition and outcome**

| Number of:                                                                                                         | NP facilitator<br>model sample<br>size | CE facilitator<br>model sample<br>size | Total<br>sample<br>size | NP facilitator<br>model<br>response rate | CE facilitator<br>model<br>response rate | Total<br>response<br>rate |
|--------------------------------------------------------------------------------------------------------------------|----------------------------------------|----------------------------------------|-------------------------|------------------------------------------|------------------------------------------|---------------------------|
| Clusters                                                                                                           |                                        |                                        |                         |                                          |                                          |                           |
| Clusters: At beginning of study                                                                                    | 17                                     | 18                                     | 35                      | n.a.                                     | n.a.                                     | n.a.                      |
| Clusters: Contributed at least one individual at baseline                                                          | 17                                     | 18                                     | 35                      | 100.0%                                   | 100.0%                                   | 100.0%                    |
| Clusters: Contributed at least one individual at first follow-up (i.e., T2 impact survey; immediate post-program)  | 17                                     | 18                                     | 35                      | 100.0%                                   | 100.0%                                   | 100.0%                    |
| Clusters: Contributed at least one individual at second follow-up (i.e., T3 impact survey; 6-months post-baseline) | 17                                     | 18                                     | 35                      | 100.0%                                   | 100.0%                                   | 100.0%                    |
| Individuals in non-attributing clusters <sup>a</sup>                                                               |                                        |                                        |                         |                                          |                                          |                           |
| Individual: At time that clusters were assigned to condition                                                       | 3091                                   | 3127                                   | 6218                    | n.a.                                     | n.a.                                     | n.a.                      |
| Individual: Who consented                                                                                          | 1787                                   | 2483                                   | 4270                    | 57.8%                                    | 79.4%                                    | 68.7%                     |
| Individual: Contributed a baseline survey                                                                          | 1455                                   | 2067                                   | 3522                    | 81.4%                                    | 83.2%                                    | 82.5%                     |
| Individual: Contributed to T2 impact survey                                                                        | 1353                                   | 1846                                   | 3199                    | 75.7%                                    | 74.3%                                    | 74.9%                     |
| Individual: Contributed to T3 impact survey                                                                        | 1180                                   | 1559                                   | 2739                    | 66.0%                                    | 62.8%                                    | 64.1%                     |
| Contributed to T3 impact survey outcomes (accounts for item nonresponse and any other analysis restrictions)       |                                        |                                        |                         |                                          |                                          |                           |
| RQ 1: Relationship Skills Knowledge                                                                                | 688                                    | 996                                    | 1684                    | 38.5%                                    | 40.1%                                    | 39.4%                     |
| RQ 2: Dating Aggression                                                                                            | 892                                    | 1253                                   | 2145                    | 49.9%                                    | 50.4%                                    | 50.2%                     |
| RQ 3: Relationship Self-Efficacy                                                                                   | 892                                    | 1265                                   | 2157                    | 49.9%                                    | 50.9%                                    | 50.5%                     |
| RQ 4: Self-Regulation                                                                                              | 918                                    | 1264                                   | 2182                    | 51.4%                                    | 50.9%                                    | 51.1%                     |

Notes: Individual-level survey response rates for both conditions and total were calculated as # individuals contributed survey / # individuals consented (rather than # individuals assigned). n.a. = not applicable.

**B. Baseline equivalence and sample characteristics**

This section describes baseline equivalence and sample characteristics of the two program conditions.

**Assessment of Baseline Equivalence.** We assessed baseline equivalence between the two program conditions, according to WWC Standards (2022) in the full analytic sample with T1 impact, nFORM Entrance, and T3 impact survey data (see Table IV.2). We also assessed baseline equivalence in each of the primary RQ analytic samples (see Appendix C, Tables C.1a, C.1c, C.1e, C.1g). Statistical tests were conducted (i.e., independent sample *t*-tests of between-group differences and Hedge's *g* effect sizes for continuous variables, and chi-square difference tests and Cox's index effect size for categorical variables) to determine whether there were differences between respondents in the NP facilitator model versus CE facilitator model. We assessed equivalence of demographic characteristics (i.e., sex, race, and relationship status) and baseline measures of all relevant outcomes for primary RQs (i.e., relationship skills knowledge, dating aggression tolerance, relationship self-efficacy, and self-regulation). These tests of group differences were completed with non-imputed data.

In the full analytic sample with T1 impact, nFORM Entrance, and T3 impact survey data, youth in the NP facilitator model and CE facilitator model conditions were equivalent on sex, as well as on baseline outcome levels of Dating Aggression Tolerance (RQ 2) and Relationship Self-Efficacy (RQ 3) (see Table IV.2). However, youth in the two conditions were not equivalent on race (with a greater proportion of non-white youth in the NP condition versus CE condition; Cox's index effect size = .89), relationship status (with a smaller proportion of youth in a relationship in the NP condition than CE condition; Cox's index effect size = .08), nor on baseline levels of Relationship Skills Knowledge outcome (RQ 1; with NP youth reporting lower knowledge than CE youth; Hedge's *g* effect size = .15), nor on baseline levels of self-regulation (RQ 4; with NP youth reporting higher self-regulation than CE youth; effect size = .08; see Table IV.2).

**Sample characteristics.** For this report, the full analytic sample of participants who completed a T1 impact survey was large (overall  $n=3,522$  individuals; NP facilitator model  $n=1,455$ ; CE facilitator model  $n=2,067$ ). Specific sample characteristics include:

- **Sex:** 51% female, 46% male
- **Race:** 51% White, 38% Black or African American, 2% Asian, 4% Other Race, 4% American Indian or Alaska Native, 1% Native Hawaiian or Other Pacific-Islander
- **Age:**  $M = 15.23$ ;  $SD = .92$ ; range from 13 to 19
- **Relationship status:** 63% not currently in a romantic relationship, 37% currently in a romantic relationship
- **Family structure:** 73% two-parent family (including original biological/adoptive parents and blended/stepfamily), 27% single-parent family

**Table IV.2 Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals completing the T1 impact, nFORM Entrance, and T3 impact surveys**

| Baseline measure                                                          | NP facilitator model mean | NP facilitator model standard deviation | CE facilitator model mean | CE facilitator model standard deviation | NP and CE difference in means | p-value of test of difference in means | Effect size (absolute value) |
|---------------------------------------------------------------------------|---------------------------|-----------------------------------------|---------------------------|-----------------------------------------|-------------------------------|----------------------------------------|------------------------------|
| Sex                                                                       |                           |                                         |                           |                                         |                               |                                        |                              |
| Female                                                                    | 55.1%                     | n.a.                                    | 55.7%                     | n.a.                                    | 0.6%                          | .78                                    | .02                          |
| Male                                                                      | 44.9%                     |                                         | 44.3%                     |                                         | 0.6%                          |                                        |                              |
| Race                                                                      |                           |                                         |                           |                                         |                               |                                        |                              |
| Other race                                                                | 66.6%                     | n.a.                                    | 31.9%                     | n.a.                                    | 34.7%                         | <.001                                  | .89^                         |
| White                                                                     | 33.4%                     |                                         | 68.1%                     |                                         | 34.7%                         |                                        |                              |
| Relationship Status                                                       |                           |                                         |                           |                                         |                               |                                        |                              |
| Dating                                                                    | 34.1%                     | n.a.                                    | 36.5%                     | n.a.                                    | 2.4%                          | .22                                    | .08^                         |
| Not in a relationship                                                     | 65.9%                     |                                         | 63.5%                     |                                         | 2.4%                          |                                        |                              |
| RQ1 Outcome: Relationship Skills Knowledge ( <i>value range 0 to 10</i> ) | 5.88                      | 1.69                                    | 6.12                      | 1.56                                    | .24                           | .003                                   | .15^                         |
| RQ2 Outcome: Dating Aggression Tolerance ( <i>value range 1 to 4</i> )    | 1.55                      | .55                                     | 1.52                      | .52                                     | .03                           | .24                                    | .05                          |
| RQ3 Outcome: Relationship Self-Efficacy ( <i>value range 1 to 4</i> )     | 3.10                      | .50                                     | 3.10                      | .50                                     | <.001                         | .97                                    | .002                         |
| RQ4 Outcome: Self-Regulation ( <i>value range 1 to 5</i> )                | 3.61                      | .58                                     | 3.56                      | .56                                     | .05                           | .05                                    | .08^                         |
| Sample size                                                               | 969                       | n.a.                                    | 1275                      | n.a.                                    | n.a.                          | n.a.                                   | n.a.                         |

Notes: n.a. = not applicable. See Table IV.3 for description of demographic covariates. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. We conducted baseline equivalence tests with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

**C. Estimation approach for primary analyses**

To estimate program impacts via equivalent effects testing, the evaluation team used a structured, multi-step approach. Before all the primary analyses steps, we first conducted preliminary analyses to confirm that model assumptions were met and to assess correlations among primary outcomes and covariates. No transformations were necessary, and outcome variables showed sufficient stability across time points. In Step 1 of our primary analyses, for each primary RQ analytic sample, we conducted initial tests of baseline equivalence between NP and CE facilitator model conditions for covariates listed in Table IV.3. Although sex and relationship status were generally equivalent across groups, differences in race and some baseline outcome measures exceeded WWC Standards (2022) thresholds (i.e., effect sizes  $> .05$ ), warranting statistical adjustment. To address these imbalances, in Step 2 we used propensity score matching for each RQ analytic sample, using race as the covariate and applying one-to-one nearest-neighbor matching with a caliper adjustment (0.2 SD) in the MatchIt package in R. Step 3 involved a secondary baseline equivalence test on the matched samples (separately for each RQ analytic sample)—revealing lingering baseline outcome imbalances for all RQ samples (1-4). Thus, to prepare for equivalent effects testing, in Step 4 we calculated adjusted predicted values for outcomes in analytic samples for RQs 1, 2, 3, and 4 (i.e., linear regression models predicting follow-up outcome by demographics, baseline outcome scores, and program condition).

In Step 5, final impact estimates were calculated using equivalent effects testing, comparing the follow-up outcomes between NP and CE program conditions using the Two One-Sided Tests (TOST) procedure (Lakens, 2017; Lakens et al., 2018), guided by a pre-specified smallest effect size of interest (SESOI) of  $d = 0.20$ . For each primary outcome, the TOST procedure evaluated whether differences between groups fell within the defined equivalence interval ( $\pm 0.20$ ), using  $p < .05$  as the threshold for statistical significance. Overall, this rigorous multi-step analytic strategy, combining propensity score matching, covariate adjustment, and equivalent effects testing, ensured a well-controlled and interpretable comparison of program effectiveness between the NP and CE facilitator model conditions. See Appendix E, Section E.1 for detailed steps of the above impact estimation steps.

**Table IV.3. Covariates included in the impact analyses**

| Covariate           | Description of the covariate                                                                                                                                                                              |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sex                 | Dichotomous variable with response options: (0) female or (1) male                                                                                                                                        |
| Race                | Dichotomous variable with response options: (0) White or (1) Other Race (i.e., Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific-Islander, other race) |
| Relationship status | Dichotomous variable with response options: (0) not currently in a relationship or (1) currently in a dating relationship                                                                                 |



## D. Implementation analyses

Refer to Appendix B, Table B.4 for detailed information on the measures, items, and composite construction used for all implementation research questions. These questions were analyzed using either frequency/percentage calculations or mean composite scores.

## E. Sensitivity analyses

Sensitivity analyses were conducted to assess the robustness of the results using non-imputed data. For analytic samples of RQ 3 and 4, we retained cases missing 20% or less items and imputed missing data. For RQ 1 and 2, we only retained cases with 100% of items. Therefore, we performed sensitivity analyses for RQ 3 and 4 using analytic samples consisting of individuals with complete data at all relevant time points.

## V. Findings

This chapter describes the key findings of the impact and implementation studies, organized by outcome. Post-hoc power analyses indicated strong power to detect equivalence across all outcomes using the TOST procedure. See Appendix E, Section E.2 for power analyses by RQ sample and Section E.3 for detailed results of the impact estimates for each primary RQ.

### A. Results of the primary impact evaluation

---

#### Key findings

---

- **RQ 1 - Relationship Skills Knowledge.** Youth in the NP facilitator model and CE facilitator model were not equivalent on relationship skills knowledge at 6-months post-baseline, with CE youth reporting significantly higher relationship skills knowledge compared to NP youth.
  - **RQ 2 - Dating Aggression Tolerance.** Youth in the NP facilitator model and CE facilitator model were equivalent on dating aggression tolerance at 6-months post-baseline.
  - **RQ 3 - Relationship Self-Efficacy.** Youth in the NP facilitator model and CE facilitator model were not equivalent on relationship self-efficacy at 6-months post-baseline, with NP youth reporting significantly higher relationship self-efficacy compared to CE youth.
  - **RQ 4 - Self-Regulation.** Youth in the NP facilitator model and CE facilitator model were equivalent on self-regulation at 6-months post-baseline.
- 

**Relationship Skills Knowledge.** NP and CE program conditions were not equivalent on relationship skills knowledge at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound = 0.381;  $p$  value test of difference in means greater or equal to upper bound <.001; NP  $M=5.991$ ,  $SD=0.919$ ,  $n=517$ ; CE  $M=6.154$ ,  $SD=0.884$ ,  $n=517$ ; see Table V.1). We conducted a follow-up independent sample  $t$ -test, comparing the adjusted outcome means at 6-months post-baseline between NP and CE conditions; the CE youth on average reported statistically significantly higher relationship skills knowledge than NP youth ( $t(1032)=-2.80$ ,  $p=.005$ ).

**Dating Aggression Tolerance.** NP and CE program conditions were equivalent on dating aggression tolerance at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound <.001;  $p$  value test of difference in means greater or equal to upper bound <.001; NP  $M=1.460$ ,  $SD=0.177$ ,  $n=697$ ; CE  $M=1.457$ ,  $SD=0.189$ ,  $n=697$ ; see Table V.1).

**Relationship Self-Efficacy.** NP and CE program conditions were not equivalent on relationship self-efficacy at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound <.001;  $p$  value test of difference in means greater or equal to upper bound =.149; NP  $M=3.125$ ,  $SD=0.178$ ,  $n=693$ ; CE  $M=3.101$ ,  $SD=0.164$ ,  $n=693$ ; see Table V.1). We conducted a follow-up

independent sample *t*-test, comparing the adjusted outcome means at 6-months post-baseline between NP and CE conditions; the NP youth on average reported statistically significantly higher relationship self-efficacy than NP youth ( $t(1304)=2.57$ ,  $p=.01$ ).

**Self-Regulation.** NP and CE program conditions were equivalent on self-regulation at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound  $<.001$ ;  $p$  value test of difference in means greater or equal to upper bound  $<.001$ ; NP  $M=3.627$ ,  $SD=0.235$ ,  $n=710$ ; CE  $M=3.630$ ,  $SD=0.240$ ,  $n=710$ ; see Table V.1).

**Table V.1. Post-intervention tests of equivalent effects using data from 6-month follow-up to address the primary research questions**

| Outcome measure                                           | NP facilitator model |       |       | CE facilitator model |       |       | Equivalent Effects Testing                   |                      |                                                                      |                                                                        |                                         |
|-----------------------------------------------------------|----------------------|-------|-------|----------------------|-------|-------|----------------------------------------------|----------------------|----------------------------------------------------------------------|------------------------------------------------------------------------|-----------------------------------------|
|                                                           | n                    | Mean  | SD    | n                    | Mean  | SD    | Smallest effect size of interest in SD units | Equivalence interval | p-value of test of difference in means lower or equal to lower bound | p-value of test of difference in means greater or equal to upper bound | Equivalent effects established (Yes/No) |
| Relationship Skills Knowledge ( <i>value range 0-10</i> ) | 517                  | 5.991 | 0.919 | 517                  | 6.154 | 0.884 | 0.20 SD                                      | (-0.180, 0.180)      | 0.381                                                                | <.001                                                                  | No                                      |
| Dating Aggression Tolerance ( <i>value range 1-4</i> )    | 697                  | 1.460 | 0.177 | 697                  | 1.457 | 0.189 | 0.20 SD                                      | (-0.037, 0.037)      | <.001                                                                | <.001                                                                  | Yes                                     |
| Relationship Self-Efficacy ( <i>value range 1-4</i> )     | 693                  | 3.125 | 0.178 | 693                  | 3.101 | 0.164 | 0.20 SD                                      | (-0.034, 0.034)      | <.001                                                                | .149                                                                   | No                                      |
| Self-Regulation ( <i>value range 1-5</i> )                | 710                  | 3.627 | 0.235 | 710                  | 3.630 | 0.240 | 0.20 SD                                      | (-0.047, 0.047)      | <.001                                                                | <.001                                                                  | Yes                                     |

Source: T3 impact surveys at 6-month follow-up

Notes: Equivalent effects were calculated using the TOST procedure. For these tests, we used the matched samples obtained via propensity score matching. For RQs 1, 2, 3, and 4, since baseline equivalence between conditions was not achieved in the propensity-score matched sample (see Tables C.1b, C.1d, C.1f, and C.1h, respectively), we used covariate-adjusted predicted T3 outcome means (predicted by the following covariates: sex, race, relationship status, baseline levels of outcome, and condition [NP or CE]). See Table III.1 for a more detailed description of each measure and Chapter IV, Section IV.C for a description of the equivalence testing approach.

## B. Results of the implementation evaluation

### Key findings

---

- **RQ 1 - Fidelity:** NP facilitator model and CE facilitator model conditions were implemented with high fidelity. On average, 91% of all curriculum activities were delivered in the NP facilitator model class series, and 93% of all curriculum activities were delivered in the CE facilitator model class series.
  - **RQ 2 & 3 - Dosage:** In the NP and CE facilitator model conditions, the large majority of participants completed 90% of programming (i.e., 91% of NP youth and 93% of CE youth completed 90% of the 12-hour class series). On average, NP youth completed 11.73 hours of programming and CE youth completed 11.76 hours of programming.
  - **RQ 4 - Quality:** Participants in NP and CE conditions rated facilitation quality of their AYRE class series very high (i.e., on scale ranging 1-5, NP participants rated on average 4.24; CE participants on average 4.29).
  - **RQ 5 - Engagement:** Participants in both NP and CE conditions rated their own engagement in the 12-hour class series moderate to high (i.e., on scale ranging 1-5, NP participants rated on average 3.66; CE participants on average 3.60).
  - **RQ 6 - Context:** Overall, the majority of NP (74%) and CE (73%) facilitator model class series occurred in participating high school Health & Wellness classes.
- 

**Fidelity.** The AYRE curriculum was implemented with high fidelity across both NP and CE conditions (i.e., 156 NP / 165 CE class series in total throughout the impact study). Out of 41 total curriculum activities (including lessons, discussions, group activities, and individual activities), NP facilitators reported teaching on average 91% of all total activities, and CE facilitators reported teaching on average 93% of all total activities across the 12-hour class series.

**Dosage.** In the analytic sample for this report (i.e., youth who completed the T1 impact, nFORM Entrance, and T3 impact surveys; NP n=969; CE n=1275), the majority of youth attended 100% of the 12-hour programming (85% of NP youth; 88% of CE youth), and an overwhelming majority attended 90% of programming, which counts as “completed” per HMRF standards (91% of NP youth; 93% of CE youth). In both NP and CE groups, 99% of youth completed at least 50% of programming, and <1% of CE youth (0% of NP youth) did not attend any classes. On average, NP youth completed 11.73 hours of programming (SD=0.82), and CE youth completed 11.76 hours of programming (SD=0.93).

**Quality:** At immediate post-program in this analytic sample, participants in both NP and CE conditions rated the quality of facilitation provided by their AYRE facilitators very highly. This included aspects such as effective teaching, providing examples, encouraging class participation, and managing time efficiently. On a 1 to 5 scale, where higher scores indicated optimal facilitation, NP participants reported an average facilitation quality score of 4.24 (SD=.79, n=797) and CE participants reported an average facilitation quality score of 4.29 (SD=.79, n=1026).

**Engagement:** Immediately after the program in this analytic sample, participants in the NP and CE conditions rated their engagement in the class series highly. This included factors such as feeling connected to other participants, actively participating in group discussions and activities, and feeling comfortable sharing their experiences. On a scale from 1 to 5, with higher scores representing greater engagement, NP participants reported an average engagement score of 3.66 (SD=.85, n=796) and CE participants reported an average engagement score of 3.60 (SD= 0.88, n=1021).

**Context:** According to program staff records, in the analytic sample for this report (NP n=969; CE n=1275) nearly three-quarters of NP students’ (74%) and CE students’ (73%) AYRE class series occurred in Health & Wellness (i.e., Health, PE) high school classrooms. The remaining AYRE class series occurred in Career and Business Education classes (i.e., Business Essentials, Career Prep, FFA; NP 5%,

CE 14%), Humanities and Social Studies (i.e., History, Social Studies, Bible; NP 3%, CE 10%), Family & Consumer Sciences (NP 3%, CE 2%), and other settings (i.e., Afterschool, Alternative school, Freshman Orientation; NP 15%, CE 1%).

## C. Results of the sensitivity analyses

---

### Key findings

---

- Complete Case Sensitivity Analyses:** The results of sensitivity equivalent effects testing using complete case (non-imputed) data for RQ 3 and RQ 4 were identical to results of primary equivalent effects testing analyses, indicating that the NP and CE conditions were not equivalent at 6-months post-baseline on relationship self-efficacy but were equivalent on self-regulation. We did not conduct sensitivity equivalent effects tests for RQ 1 and RQ 2, since we used complete case data for primary equivalent effects testing analyses.
- 

**Complete Case Sensitivity Analyses.** Sensitivity equivalent effects testing was conducted with complete case (non-imputed) data. For RQs 1 & 2, the analytic sample already included complete case data, so we did not perform sensitivity analyses on these RQs. For RQs 3 & 4, we performed sensitivity equivalent effects tests using the matched samples obtained via propensity score matching, as well as covariate-adjusted outcomes. For RQ 3, results of the sensitivity TOST equivalence tests indicated that the two program conditions were not equivalent on relationship self-efficacy at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound  $<.001$ ;  $p$  value test of difference in means greater or equal to upper bound  $=.303$ ; see Table V.2). For RQ 4, results of the sensitivity TOST equivalence tests indicated that the two program conditions were equivalent on self-regulation at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound  $=.004$ ;  $p$  value test of difference in means greater or equal to upper bound  $<.001$ ; see Table V.2).

**Table V.2. Differences in means between NP and CE facilitator model conditions estimated using alternative methods (sensitivity analyses)**

| Outcome                          | Primary approach<br>(equivalence<br>interval) | Primary approach<br>(TOST p-values) | Primary<br>approach<br>(equivalent<br>effects y/n) | Sensitivity<br>approach<br>(equivalence<br>interval) | Sensitivity<br>approach<br>(TOST p-values) | Sensitivity<br>approach<br>(equivalent<br>effects y/n) |
|----------------------------------|-----------------------------------------------|-------------------------------------|----------------------------------------------------|------------------------------------------------------|--------------------------------------------|--------------------------------------------------------|
| <b>Primary RQs</b>               |                                               |                                     |                                                    |                                                      |                                            |                                                        |
| RQ 3: Relationship Self-Efficacy | (-0.034, 0.034)                               | p1 < 0.001,<br>p2 = 0.149           | No                                                 | (-0.034, 0.034)                                      | p1 < 0.001,<br>p2 = 0.303                  | No                                                     |
| RQ 4: Self-Regulation            | (-0.047, 0.047)                               | p1 < 0.001,<br>p2 < 0.001           | Yes                                                | (-0.046, 0.046)                                      | p1 = 0.004,<br>p2 < 0.001                  | Yes                                                    |

Source: T3 impact surveys at 6-month follow-up

Notes: Sensitivity analyses were conducted with complete case (non-imputed) data. For RQs 1 & 2, the analytic sample already included complete case data; no data was imputed for these analytic samples. Equivalent effects were tested using the TOST procedure. For these tests, we used the matched samples obtained via propensity score matching. For RQs 3 & 4, since baseline equivalence between groups was not achieved in the propensity-score matched sample (see Tables C.1f & C.1h), we used covariate-adjusted predicted T3 outcome means (predicted by the following covariates: sex, race, relationship status, baseline levels of outcome, and condition [NP or CE]). See Table III.1 for a more detailed description of each measure and report section IV.C in Chapter IV for a description of the equivalence testing approach.

## **VI. Summary and conclusions**

This chapter describes the interpretation of findings, implications, considerations, and limitations of the program evaluation.

### **A. Implications**

The findings of the AYRE project provide critical insights into the effectiveness of different facilitation models for YRE and contribute to a growing body of literature underscoring the importance of relationship education in fostering youth development and resilience (e.g., Adler-Baeder et al., 2007; McElwain et al., 2017) and calling for methods to scale up access to YRE (Hawkins, 2018). Notably, the majority of studies of YRE use community educators or classroom teachers as facilitators (Hawkins, 2018). No published work has tested the comparative utility of using trained college students as YRE facilitators; therefore, this study adds novel information to the HMRE/YRE field.

The study's most significant findings center on the comparable outcomes of the "near-peer" (NP) and "community educator" (CE) facilitator models on key relational and individual outcomes, with equivalence observed on dating aggression tolerance and self-regulation six months after participating in the program. Interestingly, youth receiving the CE facilitator model reported significantly higher relationship skills knowledge at six-month follow-up, whereas youth receiving the NP facilitator model had significantly higher relationship self-efficacy. Taken together, these results underscore the viability of both NP and CE facilitator models for effective YRE delivery, as well as the unique strengths of each model. This novel test of YRE delivered by newly trained, college-aged facilitators emphasizes their effectiveness in delivering YRE programs and achieving expected outcomes mostly similar to (and in one case greater than) more experienced, seasoned YRE community educators. Our findings align with previous research emphasizing the value of relational proximity and developmental similarity in youth mentorship and education (Gehreke et al., 2024). While community educators in our study had comparatively more experience in program delivery, the near-peer educators may have held an advantage of relatability with youth, that can be associated with more student engagement. The success of the NP model, coupled with its potential for cost-efficiency and scalability, positions it as a promising alternative for expanding YRE efforts and a strategic sustainability approach.

It did appear that the CE model was slightly more effective in promoting relationship skills knowledge. This can be interpreted as added value of more facilitation experience, more advanced classroom management skills, and life experience among this group. In fact, the demographic differences of community educators compared with near-peer facilitators (e.g., more likely to be married and/or a parent) represent practical life experiences relevant to the program content which may contribute to youth's report of higher relationship skills knowledge after program participation with a community educator. These findings are consistent with studies suggesting that educator experience and familiarity with curriculum content can enhance knowledge acquisition in learners (Burroughs et al., 2019; Hawkins & Ooms, 2012).

Conversely, the NP model appeared slightly more effective in supporting youth's development of relationship self-efficacy. One possible explanation is that NP facilitators, being closer in age and life stage to participants and often conveying their own recent acquisition of better relationship skills, may be viewed as more relatable role models, which can enhance youth confidence in growing and applying relational skills. Their similar generational experiences and informal communication styles might also foster a more comfortable learning environment that encourages self-reflection and personal growth (beyond basic knowledge acquisition). Additionally, the mentoring dynamic created by NP facilitators may resonate more deeply with youth, contributing to stronger identification and internalization of key self-efficacy concepts (see Clarke-Midura et al., 2018 for parallel findings about near-peer mentors promoting mentee self-efficacy in the education field).

While we acknowledge both of these non-equivalent findings as statistically significantly and advocate for addressing these outcome differences, we also can point out that a statistically significant difference is not always meaningfully different, especially in studies with large samples. The average score at 6-months post-baseline on the relationship skills knowledge measure was 5.99 (on a scale ranging 0-10) for youth with a NP facilitator and 6.15 for youth with a CE facilitator; similarly, the average 6-month score for NP youth on the relationship self-efficacy measure was 3.13 versus CE youth average of 3.10. In practical terms these scores are quite close. Still, NP facilitators will likely benefit from future training efforts that focus on equipping them with advanced instructional techniques to further enhance youth's relationship skills knowledge acquisition, ensuring a more consistent and thorough delivery of teaching points and practices relevant to relationship skills knowledge. On the other hand, community educators may benefit from augmenting their relatable storytelling, relationship-building, and mentoring, and perhaps even including a younger near-peer as a co-facilitator for their YRE sessions. These efforts may further enhance the growth in relationship self-efficacy among youth receiving the YRE program through CE-led facilitation.

The study's implementation strengths further bolster confidence in the observed outcomes. In both facilitator models, facilitators achieved high curriculum fidelity, and the majority of participants completed over 90% of program hours and rated facilitation quality and their own engagement in the course positively. High fidelity and participant engagement suggest that the AYRE program's structure, including evidence-based curricula selections as well as comprehensive training and supervision, can effectively accommodate differing implementation contexts. This is particularly important for scaling YRE programs in states similar to Alabama, where family instability and economic challenges heighten the need for available, evidence-based interventions.

Beyond the immediate findings, this study offers broader implications for the design and delivery of school-based YRE programs. The results reaffirm the relevance of prevention science principles, particularly the role of protective factors such as experience in relationship education that can promote resilience among youth (Umberson & Karas Montez, 2010; Banyard & Cross, 2008). Additionally, the study underscores the potential of YRE programs to address critical issues, such as dating aggression and low relational self-efficacy, which have long-term implications for youth well-being and future family stability (Rice et al., 2017; Savasuk-Luxton et al., 2018).

From a policy perspective, the AYRE project demonstrates the importance of evaluating and adapting program delivery models to suit local needs and resources. A large number of youth received YRE more cost efficiently during the AYRE project (2020-2025) due to the addition of the NP model that utilized college-age students who facilitated classes for course credit, rather than payment. Policymakers and funding agencies should consider the scalability and cost-effectiveness of university-based near-peer facilitation models alongside community agency delivery of YRE, while investing in targeted training and staff support to maintain and enhance their effectiveness.

## **B. Limitations and future directions**

The AYRE project provides valuable insights into facilitation models for youth relationship education; however, several limitations must be acknowledged to contextualize the findings and guide future research. First, the quasi-experimental design used in this study, while appropriate for the real-world setting and for addressing the research questions, is less robust than randomized control trials (RCTs) in controlling for unmeasured confounding variables and ensuring comparability of groups. Although propensity score matching and covariate-adjusted outcomes were employed to minimize baseline differences, residual biases may still influence the observed outcomes (Stuart et al., 2015; Austin, 2011). As noted, community educators were more likely to have experiences with committed relationships and parenting than near-peer educators, which may be contributing to or biasing results. Future studies should aim to incorporate RCTs to establish stronger causal inferences regarding the impact of each facilitator model on youth outcomes. Second, the generalizability of findings is limited by



the study's specific context. The sample comprised youth attending high schools in Alabama, a state with unique economic and cultural characteristics that may not reflect other regions or populations. Expanding the evaluation to include more geographic, economic, and demographic groups could provide a broader understanding of the AYRE curriculum's impact using both facilitation models. Third, while the study assessed outcomes six months post-baseline, longer-term follow-up evaluations are needed to determine the sustainability of the reported outcomes, as noted in a recent YRE review (Hawkins, 2018) and meta-analysis (McElwain et al., 2017). Thus, longitudinal studies examining impacts over several years and potential spillover effects into young adult couple and family relationships would be valuable.

Future researchers could also use these two evidence-based models as a springboard for exploring other methods of YRE delivery, including a hybrid facilitation model with both a NP and CE facilitator in one classroom, which may yield innovative solutions for optimizing relatability, scalability, and effectiveness. Since NP facilitators came from a variety of discipline majors, it may also be that there was comparatively less consistency among this group than among CE facilitators. It would be interesting to examine variation in skills and effectiveness based on their backgrounds and level of knowledge and relational efficacy before and after facilitation training. This exploration could yield information relevant for enhancing the training methods for NP facilitators, emphasizing benchmarks of proficiency that best predict positive youth outcomes. Relatedly, it would be valuable to better understand the impact of the NP facilitation experience on the young adults themselves, since college age youth are part of the target population for YRE. Is there a particular advantage for their relationship skills knowledge, efficacy, and outcomes related to the teaching of YRE? To our knowledge, no study has explored or emphasized these potential concurrent benefits of YRE.

### **C. Other lessons learned**

The AYRE program evaluation provides critical insights into the implementation of youth relationship education programs and illuminates key lessons that can inform future initiatives. One significant takeaway is the robustness of the AYRE curriculum and training/supervision model. High curriculum fidelity and consistent participant engagement across both facilitation models demonstrate the adaptability and effectiveness of the AYRE curriculum content when paired with proper training and supervision. These findings suggest that a research-based curriculum, combined with tailored facilitator training and consistent oversight, can achieve favorable outcomes in differing delivery contexts.

Another valuable lesson learned was the use of young adults' feedback in our continuous quality improvement efforts. Scaling up our NP facilitation model exposed AYRE staff to a large number of college age youth over the past 4 years. Because we built in reflection exercises throughout each semester and a large-group feedback session following completion of the class series, we gained added insight into the value of the facilitation experience for young adults, stimulating our interest in the future for better documenting program impact on these older youth and the role their learning and development plays for the high school youth outcomes. From the NP facilitators we also gained added insight into the need to continually reassess program language and content in order to keep up with the ever-changing world of youth relationships. While experienced community educators have valuable feedback as well, it was the NP facilitators who were better able to point out "outdated" terminology in the lessons and to highlight relationship topics that could be enhanced or added. We were able to apply recommended language shifts and updated examples to the program without altering content. Because of the ongoing evaluation of the program and the importance of consistency across the cohorts in this study, other more substantive recommendations for program content changes and additions could be addressed with the program developer for future editions of the curriculum. These recommendations predominantly centered on the use of digital communication among dating teens. The curriculum assumes more face-to-face communication and minimally addresses the use of technology. With NP facilitator input, we adjusted some of the digital communication examples (e.g., text messages) to be more current and realistic, and added some additional discussion prompts to allow the high school youth to discuss positive and negative methods of digital communication (e.g., texting, Snapchat, DMs). These discussions always engaged the

class, indicating the topic's relevance in their lives. There were also recommendations to consider the role that peers play in healthy teen dating relationships. Currently, the program content minimally addresses this topic; therefore, this will be an area of program content we will work to enhance in the future.

Feedback also stimulated discussions regarding using more advanced instructional techniques and leveraging technology, such as videoing actual lessons and reviewing with training mentors, as well as offering more current video examples for youth in the YRE program. If videos used in the program are judged as "dated" then they can be more distracting than informing. We also found that high school youth appreciated the use of digital make-up modules used if a class was missed due to school absence or activity. This suggests the exploration in the future of using virtual booster sessions for longer-term maintenance of program effects, as well as offering asynchronous digital YRE. Expanding delivering models beyond the facilitated experience could further enhance program outreach and distribution convenience and cost-efficiency (Simpson et al., 2018).

In conclusion, the AYRE project's findings illuminate the potential of YRE programs to positively influence youth relationships and individual development using either trained college students or community educators. By applying evidence-based practices and tailoring facilitation models to local contexts, these programs can play a vital role in mitigating relational risks and fostering resilience among youth in Alabama and beyond. This research not only advances the field of YRE but also provides actionable insights for policymakers, educators, and practitioners committed to promoting healthy relationships and family stability now and in the future.

## VII. References

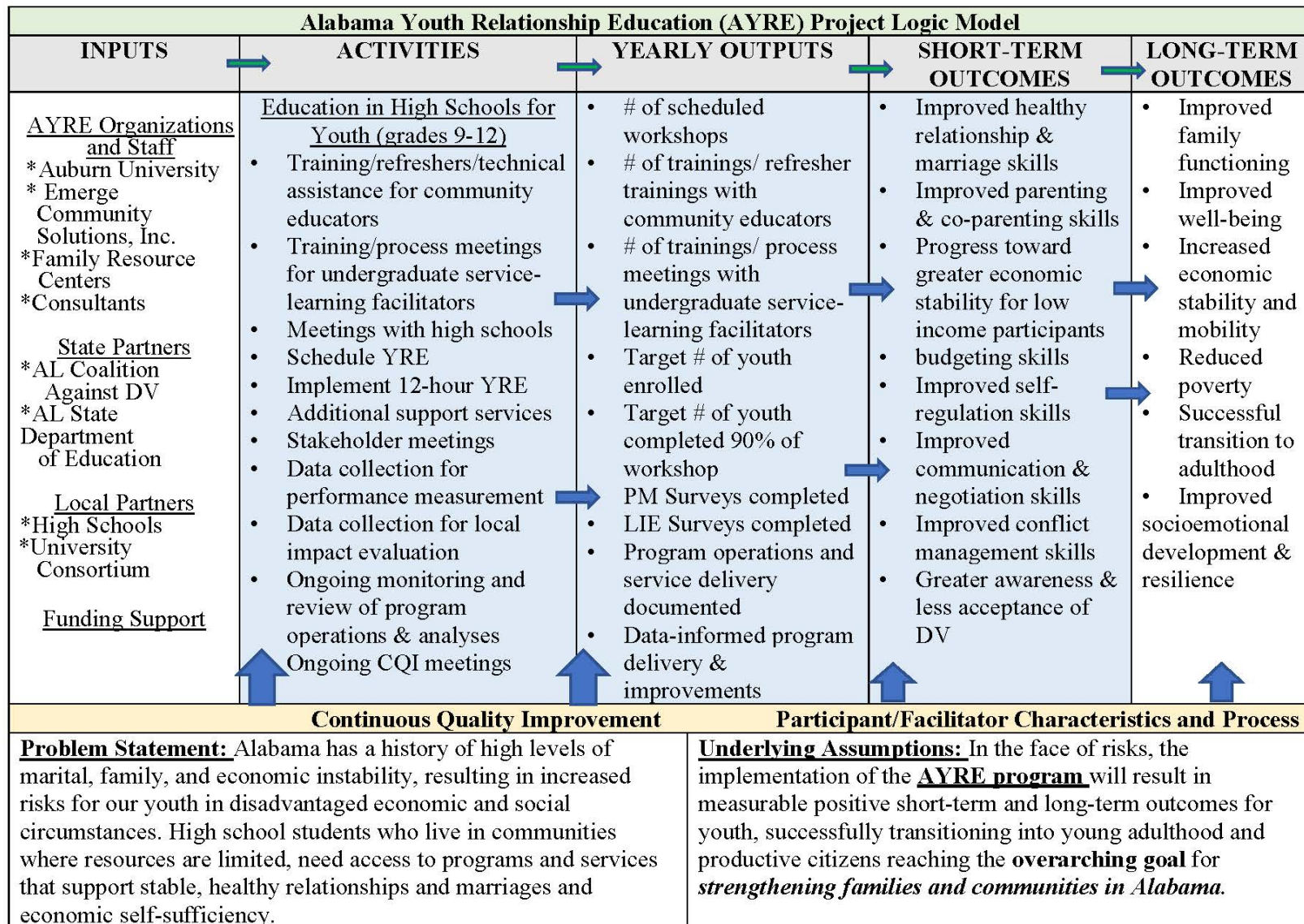
- Adler-Baeder, F., Kerpelman, J. L., Schramm, D. G., Higginbotham, B., & Paulk, A. (2007). The impact of relationship education on adolescents of diverse backgrounds. *Family Relations*, 56(3), 291-303. <https://doi.org/10.1111/j.1741-3729.2007.00460.x>
- Adler-Baeder, F., McElwain, A., McGill, J., & Gregson, K. (2025). Development and Validation of the Youth Programs-Community Education Environment Scale (YP-CEES). *Journal of Youth Development*, 20(1), 13. <https://open.clemson.edu/jyd/vol20/iss1/13>
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavior Research*, 46(3), 399-424. <https://doi.org/10.1080/00273171.2011.568786>
- Banyard, V. L., & Cross, C. (2008). Consequences of teen dating violence: Understanding intervening variables in ecological context. *Violence Against Women*, 14(9), 998-1013. <https://doi.org/10.1177/1077801208322058>
- Bouchey, H. A., & Furman, W. (2003). Dating and romantic experiences in adolescence. In G. R. Adams & M. D. Berzonsky (Eds.), *Blackwell handbook of adolescence* (pp. 313-329). Blackwell Publishing. <https://doi.org/10.1002/9780470756607.ch15>
- Bradford, A. B., Adler-Baeder, F., Ketring, S. A., & Smith, T. A. (2012). The role of participant-facilitator demographic match in couple and relationship education. *Family Relations*, 61(1), 51-64. <https://doi.org/10.1111/j.1741-3729.2011.00679.x>
- Burroughs, N., Gardner, J., Lee, Y., Guo, S., Touitou, I., Jansen, K., & Schmidt, W. (2019). A review of the literature on teacher effectiveness and student outcomes. In N. Burroughs, J. Gardner, Y. Lee, S. Guo, I. Touitou, K. Jansen, & W. Schmidt (Eds.), *Teaching for excellence and equity: Analyzing teacher characteristics, behaviors and student outcomes with TIMSS* (pp. 7-17). Springer International Publishing. [https://doi.org/10.1007/978-3-030-16151-4\\_2](https://doi.org/10.1007/978-3-030-16151-4_2)
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Clarke-Midura, J., Poole, F., Pantic, K., Hamilton, M., Sun, C., & Allan, V. (2018, February). How near peer mentoring affects middle school mentees. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (pp. 664-669). <https://doi.org/10.1145/3159450.3159525>
- Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., ... & Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*, 48(10), 1013-1022. <https://doi.org/10.1037/0003-066X.48.10.1013>
- Curtis, C. & Stolzenbach, C. (2017). *Mind matters*. The Dibble Institute. <https://dibbleinstitute.org/our-programs/mind-matters-overcoming-adversity-and-building-resilience/>

- Friend, Daniel, Rebecca Piatt, Avery Hennigar, Armando Yañez, Brandon Hollie, Sonia Alves, and Angela D'Angelo. (2024). *Healthy marriage and relationship education program evaluation toolkit* (Working toolkit). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Futris, T. G., Sutton, T. E., & Duncan, J. C. (2017). Factors associated with romantic relationship self-efficacy following youth-focused relationship education. *Family Relations*, 66, 777-793. <https://doi.org/10.1111/fare.12288>
- Gehreke, L., Schilling, H., & Kauffeld, S. (2024). Effectiveness of peer mentoring in the study entry phase: A systematic review. *Review of Education*, 12(1), e3462. <https://doi.org/10.1002/rev3.3462>
- Graham, S. E., & Kurlaender, M. (2011). Using propensity scores in educational research: General principles and practical applications. *The Journal of Educational Research*, 104, 340-353. <https://doi.org/10.1080/00220671.2010.486082>
- Harris, H., & Horst, S. J. (2016). A brief guide to decisions at each step of the propensity score matching process. *Practical Assessment, Research, & Evaluation*, 21(1), 4. <https://doi.org/10.7275/yq7r-4820>
- Hawkins, A. J. (2018). Shifting the relationship education field to prioritize youth relationship education. *Journal of Couple & Relationship Therapy*, 17(3), 165-180. <https://doi.org/10.1080/15332691.2017.1341355>
- Hawkins, A. J., & Ooms, T. (2012). Can marriage and relationship education be an effective policy tool to help low-income couples form and sustain healthy marriages and relationships? A review of lessons learned. *Marriage & Family Review*, 48, 524-554. <https://doi.org/10.1080/01494929.2012.677751>
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1-28. <https://doi.org/10.18637/jss.v042.i08>
- Julious, S. A. (2004). Sample sizes for clinical trials with normal data. *Statistics in Medicine*, 23(12), 1921-1986. <https://doi.org/10.1002/sim.1783>
- Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychology and Personality Science*, 8(4), 355-362. <https://doi.org/10.1177/1948550617697177>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269. <https://doi.org/10.1177/2515245918770963>
- McElwain, A., McGill, J., & Savasuk-Luxton, R. (2017). Youth relationship education: A meta-analysis. *Children and Youth Services Review*, 82, 499-507. <https://doi.org/10.1016/j.childyouth.2017.09.036>
- Mendo-Lázaro, S., León-del-Barco, B., Polo-del-Río, M. I., Yuste-Tosina, R., & López-Ramos, V. M. (2019). The role of parental acceptance–rejection in emotional instability during adolescence.

- International Journal of Environmental Research and Public Health*, 16(7), Article 1194.  
<https://doi.org/10.3390/ijerph16071194>
- Pearson, M. (2018). *Relationship Smarts PLUS 4.0*. The Dibble Institute. <https://dibbleinstitute.org/our-programs/relationship-smarts-plus-5-0/>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Rice, T., McGill, J., & Adler-Baeder, F. (2017). Relationship education for youth in high school: preliminary evidence from a non-controlled study on dating behavior and parent-adolescent relationships. *Youth Care Forum*, 46, 51-68. <https://doi.org/10.1007/s10566-016-9368-8>
- Savasuk-Luxton, R., Adler-Baeder, F., Haselschwerdt, M.L. (2018). Understanding change in violence related attitudes for adolescents in relationship education. *Journal of Adolescence*, 63, 153-164. <https://doi.org/10.1016/j.adolescence.2017.12.012>
- Simpson, D. M., Leonhardt, N. D., & Hawkins, A. J. (2018). Learning about love: A meta-analytic study of individually-oriented relationship education programs for adolescents and emerging adults. *Journal of Youth and Adolescence*, 47(3), 477-489. <https://doi.org/10.1007/s10964-017-0725-1>
- Solomon, S. (2009). *Money habitudes: A guide for professionals working with money-related issues* (3rd ed.). LifeWise Productions.
- Stuart, E. A., Bradshaw, C. P., & Leaf, P. J. (2015). Assessing the generalizability of randomized trial results to target populations. *Prevention Science*, 16(3), 475-485. <https://doi.org/10.1007/s11121-014-0513-z>
- Umberson, D., & Karas Montez, J. (2010). Social relationships and health: A flashpoint for health policy. *Journal of Health and Social Behavior*, 51(1\_suppl), S54-S66. <https://doi.org/10.1177/0022146510383501>
- Van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45, 1-67. <https://doi.org/10.18637/jss.v045.i03>
- Weiss, H. B., & Jacobs, F. H. (Eds.). (2008). *Evaluating family programs: Current issues in theory and policy*. Transaction Publishers.
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook* (Version 5.0). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://ies.ed.gov/ncee/wwc/Handbooks>
- Wood, R. G., Moore, Q., Clarkwest, A., & Killewald, A. (2014). The long-term effects of Building Strong Families: A program for unmarried parents. *Journal of Marriage and Family*, 76, 446-463. <https://doi.org/10.1111/jomf.12094>

## **Appendix Tables and Figures to Supplement Final Impact Report**

## Appendix A: Logic Model





## Appendix B: Study Design, Data Collection, and Analytic Sample

This appendix gives details about recruitment and enrollment into the study, the consent and data collection process (for both impact and implementation RQs), outcome measures (for both impact and implementation RQs), statistical comparison of facilitator demographics, and CONSORT diagrams (at the individual and cluster levels).

### B.1. QED Recruitment & Enrollment

Recruitment and enrollment procedures were identical for both NP and CE facilitator models. All nine program sites were responsible for recruiting a portion of the evaluation sample. Both NP and CE program sites established individual relationships with teachers of subjects such as Health & Wellness, Business & Career Planning, Family & Consumer Science in partnering high schools (see below in this section for more details on school partnerships). The program sites then coordinated with partnering classroom teachers to determine schedule of delivery for each semester (see above Chapter II, Section II.B Dosage and implementation schedule). For both conditions, youth participants who were enrolled in the specific high school classes with a partnering teacher were offered the chance to enroll into our impact study (via parent consent and youth assent) in 20 separate cohorts, which were named by the month and year in which their class series began: Jan 2022, Feb 2022, Mar 2022, Apr 2022, Aug 2022, Sep 2022, Oct 2022, Nov 2022, Jan 2023, Feb 2023, Mar 2023, Apr 2023, Aug 2023, Sep 2023, Oct 2023, Nov 2023, Jan 2024, Feb 2024, Mar 2024, and Apr 2024.

Since the impact study used a QED with cluster assignment at the high school level (based on location and proximity to program site), efforts were made during planning, startup, and impact study periods to promote and monitor baseline equivalence of the NP and CE groups. NP and CE program sites served non-overlapping counties in Alabama, and close attention was paid throughout the grant cycle to the urban/rural setting of counties served and participant demographics of partnering high schools, to promote demographic balance between study conditions. At the start of this grant cycle, NP and CE program sites had a balance in number of high school partnerships in counties with similar settings (i.e., approximately 1/3 urban counties and 2/3 rural counties for each NP facilitator model program sites and CE facilitator model program sites). Similarly, at the start of the grant cycle, NP and CE program sites had a balance in youth demographics at partnering high schools (i.e., serving approximately 40% non-white students across high schools in each NP and CE facilitator models). During the planning period, additional high schools were selected for each NP and CE group based on their similarity of county setting (rural/urban) and student demographics. Throughout the first year of startup programming (i.e., March 2021 – Dec 2021), the evaluation team monitored baseline numbers and demographics of student participants and assessed for comparability across the two groups. We made adjustments to high schools and numbers of students targeted throughout that first startup year with the aim of balancing youth participant demographics.

After the startup year, throughout the rest of the five-year grant cycle both CE and NP program sites continued to initiate new partnerships with local high schools in order to meet target enrollment numbers and spread the reach of AYRE programming. In order to meet the high target numbers of the AYRE impact study, NP and CE program sites established new partnerships within their own county and the surrounding counties, particularly if a school partner chose not to continue (typically for scheduling reasons). NP and CE program sites made sure to not overlap with counties served by the other group.

Throughout the impact study data collection period (Jan 2022 – Apr 2024), the evaluation team assessed differences in baseline demographic variables (e.g., sex, race, relationship status) across the two program conditions on a quarterly basis. Over the course of the 5-year grant cycle, most school partnerships remained constant; however, some new high school partnerships were developed. The evaluation team noted that some differences between the counties and high schools served by the two



conditions arose, with NP program sites serving a higher proportion of urban counties and high schools with a greater percentage of non-white students than CE program sites. We closely monitored this balance throughout the impact study period and continued to make best efforts to both reach program enrollment targets and to balance study conditions. As this was a QED, baseline differences in participants by study condition were balanced using analytic methods (i.e., propensity score matching, see Chapter IV, Section IV.C. below).

Participating high schools for the NP facilitator model (n = 17) included: Auburn, Beauregard, Beulah, Booker T Washington, Brew Tech, Carver, Glenwood, Lanett, Loachapoka, Notasulga, Opelika Learning Center, Park Crossing, Percy L Julian, Reeltown, Tallassee, Valiant Cross Academy, and Valley. Participating high schools for the CE facilitator model (n = 18) included: Elmore County, Escambia County, Etowah, Gadsden City, Gaston, Genesis Alternative, Glencoe, Hillcrest, Hokes Bluff, Holtville, Mountain View Medical Center, New Heights, Sardis, Southside, Stanhope Elmore, Sylacauga, West End, and Wetumpka.

AYRE's long-standing, established community partnerships with both partner sites and local high schools were critical to this program's success. At the beginning of this 2020-2025 ACF grant cycle, AYRE had many years of partnership with all program sites, except for one of the NP program sites which was a newly established partnership for the purpose of overseeing NP facilitators for this particular impact study. As a result of prior HMRF grant cycles, all AYRE program site partners were experienced in delivering YRE programming in high schools using the CE facilitator model. AYRE's overall approach was collaborative and built using engaged participatory methods (Small & Uttal, 2005). AU AYRE staff and program site staff stayed visible in the community and were strategic in maintaining ongoing communication with school personnel throughout the year. This included planning for program delivery and evaluation well in advance (e.g., meeting during academic fall semester to plan for implementation during academic spring semester). Program site staff worked cooperatively with key school personnel to reach an agreed upon general timeline (e.g., data collection dates, program delivery dates, class time allotted for programming) and worked around school activities and events (e.g., school holidays, testing). This collaboration included noting the school personnel's preferred method for communication to ensure open, ongoing communication and asking for input on best methods for arranging a safe, welcoming program delivery and evaluation environment.

## **B.2. Consent Process**

The consent process was identical for participants in both conditions. Program staff first worked with school staff to receive class rosters for students who were enrolled in participating classes. Once class rosters were obtained, program staff entered the student information in the nFORM system using the Application Form. Before the workshop series began, program staff went to the school and used a script to inform youth about the informed consent/assent letter. They highlighted the evaluation of the AYRE program, the confidentiality and use of data collected, and the voluntary nature of participation. The staff then asked for affirmation that the students understood about the informed consent/assent and evaluation study process, and staff emphasized that if youth or their parents chose for them not to take part in the evaluation, they could still be allowed to participate in the AYRE workshop series. The IRB-approved consent/assent letter was distributed to youth as a hard-copy (to take home for a parent/guardian signature) or emailed to parents/guardians as a secure Qualtrics link where parents and youth could provide consent/assent electronically. Classroom teachers assisted in obtaining consent by distributing the hard-copy informed consent letter or electronic form to students and their parents/guardians. Only students who returned signed consent/assent letters (via paper copies or Qualtrics links) completed nFORM and AYRE impact surveys. Any paper copies of the signed informed consents were collected and put in a secure place (i.e., locked cabinet) per the approved IRB protocol.

Consented participants' contact information was uploaded into a sophisticated, secure participant tracking system, via Ripple Science online software (<https://www.ripplescience.com>). The evaluation team

managed the Ripple Science online databases linking identifying information (i.e., names, addresses, phone numbers) to unique individual IDs (created by nFORM). Access to Ripple databases was limited to only trained, eligible staff.

### **B.3. Data Collection Detailed**

On “baseline data collection day”, partner staff used a script to explain the nFORM and AYRE impact surveys to be completed on the provided tablet, the ACASI system, and the tablet functionality. Program staff generated passcodes for the nFORM Applicant Characteristics form, wrote the passcodes on an index card with the nFORM participant ID (no name), and provided this card to the student. The student used the nFORM ID and survey passcode to complete the nFORM Applicant Characteristics form and then notified the program staff member when they had completed the first survey. Using the same nFORM ID, the student then completed the T1 impact survey (using a link provided by the staff member). Finally, the staff member generated an nFORM Entrance survey passcode for the student and wrote it on the index card, and the student used this ID and passcode to complete the nFORM Entrance survey. Upon completion of all three surveys, the student returned the tablet and notecard to the AYRE staff member. The staff member checked that the system indicated the nFORM Entrance survey was complete. Procedures for audio-assistance and privacy protection were followed according to the IRB protocol.

On “immediate post-program data collection day”, program staff again provided students with their nFORM ID on an index card (no name), and students used this to complete the T2 impact survey on a provided tablet via an electronic link. Partner staff then generated an nFORM survey passcode and wrote it on the index card with their participant ID. The student used the ID and passcode to complete the nFORM Exit survey. Upon completion of both surveys, the student returned the tablet and index card.

Although all high schools had Wi-Fi available, program staff brought back-up hotspots to schools on data collection days or coordinated in advance to schedule access to a computer lab for data collection. In the rare case that students had significant problems with the internet (e.g., power or Wi-Fi outage), program staff were prepared to provide paper versions of surveys. Program staff overnight mailed paper surveys to the AU program team, who managed data entry of paper surveys into nFORM surveys and AYRE Qualtrics impact surveys (per IRB guidelines and approval from ACF).

The local evaluation team conducted the final T3 impact survey data collection. Using the Ripple participant management software (i.e., an advanced participant data collection tracking system that meets IRB privacy requirements), youth participants and their parents were emailed and texted one week prior to sending out the T3 impact survey, asking for updates to contact information. Contact information was accordingly updated in Ripple, and the following week the evaluation team texted and emailed electronic survey links to participants. Over the remaining 4-week survey window period, the evaluation team sent text and email completion reminders to participants who had not yet finished the survey. Evaluation team members tracked participant contact and Qualtrics survey completion using Ripple software, and they generated survey compensation lists by downloading survey completion from Ripple. AU program staff then emailed a \$25 electronic gift card to youth as compensation for T3 impact survey completion.

**Table B.1. Key features of data collection for the impact analysis**

| Study group                | Data source                             | Timing of data collection                          | Mode of data collection                                                                                                                                                                        | Parties responsible for data collection | Start and end date of data collection |
|----------------------------|-----------------------------------------|----------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|---------------------------------------|
| NP & CE facilitator models | T1 (baseline) impact survey (Qualtrics) | Baseline (pre-program before class sessions begin) | T1 impact survey was completed in-person, via link to Qualtrics survey and IDs/passcodes provided to students in high school classrooms on tablets provided by program staff                   | Program staff                           | January 2022 through May 2024         |
| NP & CE facilitator models | nFORM entrance survey                   | Baseline (pre-program before class sessions begin) | nFORM entrance survey was completed in-person, via link to nFORM online platform and IDs/passcodes provided to students in high school classrooms on tablets provided by program staff         | Program staff                           | January 2022 through May 2024         |
| NP & CE facilitator models | T3 (6-month) impact survey (Qualtrics)  | 6-months post-baseline                             | T3 impact survey was completed on participant's own device via an individualized Qualtrics survey link emailed & texted to participants; participant IDs were embedded in individualized links | Evaluation staff                        | July 2022 through November 2024       |

Notes: NP=near-peer. CE=community educator. T1=Time 1 survey (i.e., baseline). T3=Time 3 follow-up survey (i.e., 6-months post-baseline).

**Table B.2. Key features of data collection for the implementation analysis**

| Implementation element | Research question                                                                                                                                                                                            | Data source                                                                                                             | Timing and frequency of data collection          | Party responsible for data collection |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|---------------------------------------|
| Fidelity               | 1. For each condition, according to facilitator reports after each class session, what average percentage of curriculum content was taught?                                                                  | Paper facilitator fidelity checklists                                                                                   | After each class session in each workshop series | Program staff; evaluation staff       |
| Dosage                 | 2. For each condition, what percentage of individuals attended: 0%, 50%, 90% or 100% of the 12-hour workshop series?<br><br>3. For each condition, what is the average number of programming hours received? | Workshop session attendance records in nFORM                                                                            | After each class session in each workshop series | Program staff; evaluation staff       |
| Quality                | 4. For each condition, what is the average participant post-program rating of facilitation quality?                                                                                                          | T2 (immediate post-program) impact surveys: 5-item facilitation quality subscale of YPCEES (Adler-Baeder et al., 2025)  | Immediate post-program                           | Program staff; evaluation staff       |
| Engagement             | 5. For each condition, what is the average participant post-program rating of self-engagement during the class series?                                                                                       | T2 (immediate post-program) impact surveys: 3-item individual engagement subscale of YPCEES (Adler-Baeder et al., 2025) | Immediate post-program                           | Program staff; evaluation staff       |
| Context                | 6. For each condition, in what type of high school classes were the AYRE workshop series delivered (by percentage)?                                                                                          | Program staff records                                                                                                   | Throughout impact study period                   | Program staff                         |

Notes: T2=Time 2 follow-up survey (i.e., immediate post-program).

**Table B.3. Outcome measures, items, value scoring, and description of composite calculation for all primary outcomes**

| RQ | Outcome Measure               | Items                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Values                    | Description of Composite                                                                                                                |
|----|-------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Relationship Skills Knowledge | <ol style="list-style-type: none"> <li>1. Your day-to-day decisions, the negative, and positive influences in your life, and your relationships will shape your: <ol style="list-style-type: none"> <li>a. 1 = Possible selves (correct answer)</li> <li>b. 2 = Multiple personalities</li> <li>c. 3 = Maturity</li> <li>d. 4 = Romantic attractions</li> </ol> </li> <li>2. Which of these is one of the 4 Dimensions of Maturity? <ol style="list-style-type: none"> <li>a. 1 = Financial</li> <li>b. 2 = Emotional (correct answer)</li> <li>c. 3 = Philosophical</li> <li>d. 4 = Spiritual</li> </ol> </li> <li>3. According to the Relationship Pyramid, which of these is a smart foundation on which to build a relationship? <ol style="list-style-type: none"> <li>a. 1 = Positive starters (correct answer)</li> <li>b. 2 = Mature love</li> <li>c. 3 = Deepening and Developing</li> <li>d. 4 = Sexual Attraction</li> </ol> </li> <li>4. Which of these is one of the Seven Principles of Smart Relationships? <ol style="list-style-type: none"> <li>a. 1 = Expect to agree about everything</li> <li>b. 2 = Expect good communication (correct answer)</li> <li>c. 3 = Opposites attract</li> <li>d. 4 = Hide your true feelings</li> </ol> </li> <li>5. Which of these is NOT one of the basic human needs? <ol style="list-style-type: none"> <li>a. 1 = Physical</li> <li>b. 2 = Emotional</li> <li>c. 3 = Social</li> <li>d. 4 = Philosophical (correct answer)</li> </ol> </li> <li>6. Which of these steps is part of the Low-Risk, Deciding Approach to relationships? <ol style="list-style-type: none"> <li>a. 1 = Fall in love quickly</li> <li>b. 2 = Learn more about each other (correct answer)</li> <li>c. 3 = Be physically intimate right away</li> <li>d. 4 = Ignore problems</li> </ol> </li> <li>7. Which of these is a “red flag” warning sign of abuse in relationships? <ol style="list-style-type: none"> <li>a. 1 = Calming down before solving a conflict.</li> <li>b. 2 = Having disagreements.</li> <li>c. 3 = Getting jealous when you spend time with other people. (correct answer)</li> <li>d. 4 = Teasing each other when you're both joking and having fun.</li> </ol> </li> </ol> | 1=Correct,<br>0=Incorrect | Sum composite of correct answers on all 10 items (value range 0-10); higher scores represent more knowledge about healthy relationships |

| RQ | Outcome Measure             | Items                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Values                                               | Description of Composite                                                                                        |
|----|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
|    |                             | <p>8. Which of these is one of the rules or steps for taking a Time Out?</p> <p>a. 1 = Tell your partner they need a time out.</p> <p>b. 2 = Do or say something to calm yourself. (correct answer)</p> <p>c. 3 = Keep thinking about what made you angry.</p> <p>d. 4 = Walk away and avoid talking about the problem.</p> <p>9. What is the WWA formula?</p> <p>a. 1 = Who, What, Alternative</p> <p>b. 2 = What, When or Where, Affect (correct answer)</p> <p>c. 3 = Who, Watch and Wait, Aware</p> <p>d. 4 = What, Willingness, Ask Permission</p> <p>10. Which of these is an example of a healthy relationship?</p> <p>a. 1 = You feel like you have to look a certain way in order to keep their attention or love.</p> <p>b. 2 = One partner needs to be the "boss."</p> <p>c. 3 = This relationship is based almost exclusively on sex.</p> <p>d. 4 = This couple has chemistry, but they also enjoy talking and getting to know each other. (correct answer)</p> |                                                      |                                                                                                                 |
| 2  | Dating Aggression Tolerance | <p>1. A person who makes their partner angry on purpose deserves to be hit.</p> <p>2. Sometimes physical violence, such as hitting or pushing, is the only way to express your feelings.</p> <p>3. Violence between dating partners is a personal matter, and people should not interfere.</p> <p>4. It's okay to stay in a relationship even if you're afraid of your boyfriend/girlfriend.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 1 = Strongly Disagree to 4 = Strongly Agree          | Mean composite of all 4 items (value range 1-4); higher scores indicate more tolerance toward dating aggression |
| 3  | Relationship Self-Efficacy  | <p>1. I believe I will be able to effectively deal with conflicts that arise in my relationship.</p> <p>2. I feel good about my ability to make a romantic relationship last.</p> <p>3. I am very confident when I think of having a stable, long-term relationship.</p> <p>4. I have the skills needed for a lasting, stable, long-term relationship.</p> <p>5. I am able to recognize the warning signs of a bad relationship.</p> <p>6. I know what to do when I recognize the warning signs of a bad relationship.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 1 = Strongly Disagree to 4 = Strongly Agree          | Mean composite of all 6 items (value range 1-4); higher scores indicate greater relationship self-efficacy      |
| 4  | Self-Regulation             | <p><b>Adolescent Self-Regulatory Inventory:</b></p> <p>1. When I'm sad, I can usually start doing something that will make me feel better.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 1 = Not at all true for me to 5 = Really true for me | Mean composite of all 10 items (value range 1-                                                                  |

| RQ | Outcome Measure | Items                                                                                                                                                                                                                                                                                      | Values | Description of Composite                           |
|----|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|----------------------------------------------------|
|    |                 | 2. I can stop myself from doing things like throwing objects when I'm mad.<br>3. I can calm myself down when I'm excited or all wound up.<br>4. When I have a serious disagreement with someone, I can talk calmly about it without losing control.                                        |        | 5); higher scores indicate greater self-regulation |
|    |                 | <b>Social-Emotional Health Survey-Secondary:</b>                                                                                                                                                                                                                                           |        |                                                    |
|    |                 | 1. I feel bad when someone gets his or her feelings hurt.<br>2. I try to understand what other people go through.<br>3. I try to understand how other people feel and think.<br>4. I can wait for what I want.<br>5. I don't bother others when they are busy.<br>6. I think before I act. |        |                                                    |

Notes: RQ=research question.

**Table B.4. Implementation RQ measure items and values/scoring**

| RQ            | Implementation Measure Items & Values/Scoring                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fidelity #1   | <ul style="list-style-type: none"> <li>• <b>Please check each activity you completed of the Relationship Smarts Plus 4.0 curriculum:</b> 1.1 Possible Selves Tree (How I See Myself, Creating My Possible Selves Tree, Pair/Share Time; 1.2 My Friends and My Future; 2.1 The Four Dimensions of Maturity; 2.3 Clarifying Values (Values Auction &amp; Discussion); 3.1 Test Your Love Smarts; 3.2 Relationship Pyramid (Group Brainstorm, Build a Relationship, Inverted Pyramid); 3.3 The Chemistry of Attraction; 4.1 Smart or Not-So-Smart?; 4.2 Seven Principles of Smart (Identify that Principle); 5.1 How Can You Tell (3 Questions to Ask, Relationship Sculptures, Share &amp; Discussion); 5.2 Unmet Needs &amp; Unhealthy; 5.3 Having Fun (Fun Brainstorm); 6.3 Types of IPV and Dating Violence; 6.2 Early Warnings &amp; Red Flags (Red Flag Competition); 6.1 Breaking Up (Stay Together or Break Up?); 7.1 Healthy Selves (Sorting Baggage); 7.2 High-Risk Sliding Approach; 7.4 The Low-Risk Deciding Approach; 8.3 Angry Brains &amp; Time Outs (Lecture Doodle Video); 8.4 Speaker Listener Technique; 9.1 Raise Issues Effectively (Ineffective Ways to Complain, WWA, Good or Bad Complaint?); 9.2 Hidden Issues; 9.3 Problem-Solving Model; 13.1 What's Real?; 13.2 Constantly Connected...; 13.3 Romantic Relationships &amp; Social Media (Share Social Media Safety Resource)</li> <li>• <b>Please check each activity you completed of the Mind Matters curriculum:</b> 1.1 Introduction to Mind Matters; 1.3 Focused Breathing; 1.5 5-4-3-2-1 Skill (Importance of Practice); 2.1 Body Scan; 2.2 Key Content (The Role of Emotions, What is Emotional Intelligence?, Benefits of Emotional Intelligence); 4.1 Building Empathy (Building Empathy Exercise); 4.2 Loving Kindness; 5.1 Mapping Your Support; 6.1 Introduction of Trauma; 6.3 Effects of Trauma on...; 6.5 Hope: Methods to Build Your Brain; 11.1 Code of Honor (Code of Honor Worksheet, Honor Shield); 11.2 Asking for Help (Rank Your Own Fears)</li> <li>• <b>Please check each activity you completed on Erin's Law:</b> Erin's Law (Background of Erin's Law, Share Erin's Law Resource)</li> <li>• <b>Please check each activity you completed of the Money Habitudes curriculum:</b> Money Habitudes Card Sort, Interpretation, Wrap-Up</li> </ul> <p><b>Values/Scoring:</b> All 41 session activities were marked for completion (1=Yes, 0=No)</p> |
| Dosage #2     | <p>Individual attendance was marked after each of the workshop sessions</p> <p><b>Values/Scoring:</b> 0=0 sessions, 1=at least 1 session, 2=50% or more of sessions, 3=all 6 sessions</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Quality #3    | <p>Facilitation Quality subscale of the YPCEES (Adler-Baeder et al., 2025)</p> <ol style="list-style-type: none"> <li>1. The facilitators explained the course material clearly.</li> <li>2. The facilitators effectively encouraged class participation.</li> <li>3. The facilitators managed classroom comments and behavior appropriately.</li> <li>4. The facilitators managed the class time well.</li> <li>5. The facilitators shared appropriate examples and experiences to help us understand the course material.</li> </ol> <p><b>Values/Scoring:</b> 1=strongly disagree, 2=disagree, 3=neither/mixed, 4=agree, 5=strongly agree</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Engagement #4 | <p>Individual Engagement subscale of the YPCEES (Adler-Baeder et al., 2025)</p> <ol style="list-style-type: none"> <li>1. I felt connected to other participants in the class.</li> <li>2. I felt comfortable sharing experiences with the group.</li> <li>3. I was actively engaged in group discussions and activities.</li> </ol> <p><b>Values/Scoring:</b> 1=strongly disagree, 2=disagree, 3=neither/mixed, 4=agree, 5=strongly agree</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Context #5    | <p>In what type of high school class was this AYRE workshop series delivered?</p> <p><b>Values/Scoring:</b> 1=Career &amp; Business Education (Business Essentials, Career Prep, FFA), 2=Family &amp; Consumer Sciences, 3=Health &amp; Wellness (Health, PE), 4=Humanities &amp; Social Studies (History, Social Studies, Bible), 5=Other (Afterschool, Alternative School, Freshman Orientation)</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

Notes: RQ=research question.



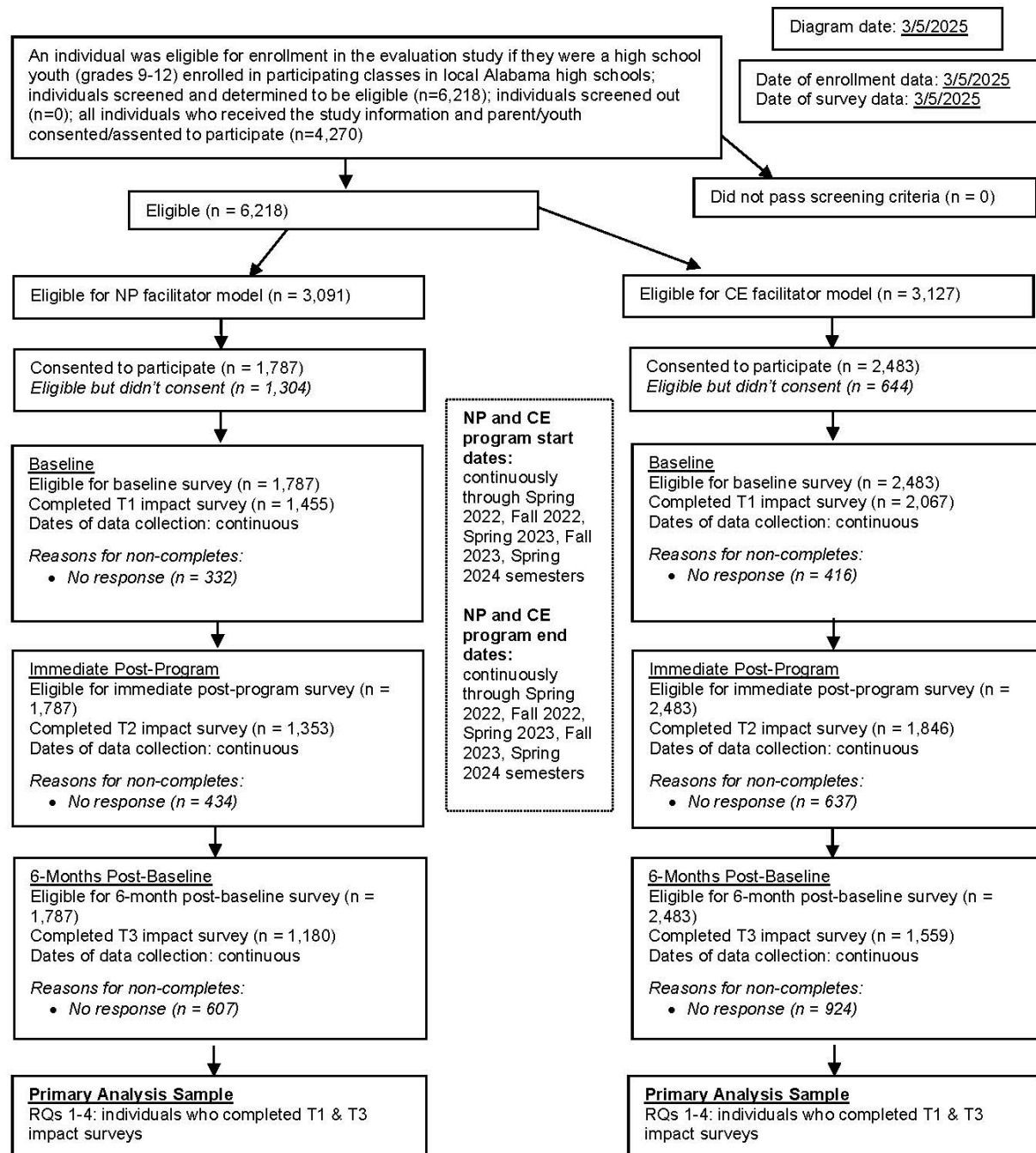
**Table B.5 Comparison of Facilitator Demographics by Condition: NP vs. CE**

| Baseline measure        | NP facilitator model mean or % | NP facilitator model standard deviation | CE facilitator model mean or % | CE facilitator model standard deviation | NP and CE difference in means or %s | p-value of test of difference in means or %s | Effect size (absolute value) |
|-------------------------|--------------------------------|-----------------------------------------|--------------------------------|-----------------------------------------|-------------------------------------|----------------------------------------------|------------------------------|
| Sex                     | 82.9%                          | n.a.                                    | 71.0%                          | n.a.                                    | 11.9%                               | .16                                          | .42^                         |
| Female                  | 17.1%                          |                                         | 29.0%                          |                                         |                                     |                                              |                              |
| Male                    |                                |                                         |                                |                                         |                                     |                                              |                              |
| Race                    | 64.6%                          | n.a.                                    | 64.5%                          | n.a.                                    | 0.1%                                | .99                                          | .00                          |
| Other race              | 35.4%                          |                                         | 35.5%                          |                                         |                                     |                                              |                              |
| White                   |                                |                                         |                                |                                         |                                     |                                              |                              |
| Age                     | 22.41                          | 5.22                                    | 43.30                          | 14.76                                   | 20.89                               | <.001                                        | 2.36^                        |
| Education               | 78.0%                          | n.a.                                    | 19.4%                          | n.a.                                    | 58.6%                               | <.001                                        | 1.65^                        |
| No 4-year degree        | 22.0%                          |                                         | 80.6%                          |                                         |                                     |                                              |                              |
| 4-year degree or higher |                                |                                         |                                |                                         |                                     |                                              |                              |
| Married Status          | 92.7%                          | n.a.                                    | 25.8%                          | n.a.                                    | 66.9%                               | <.001                                        | 2.20^                        |
| Not married             | 7.3%                           |                                         | 74.2%                          |                                         |                                     |                                              |                              |
| Married                 |                                |                                         |                                |                                         |                                     |                                              |                              |
| Parent Status           | 91.5%                          | n.a.                                    | 25.8%                          | n.a.                                    | 65.7%                               | <.001                                        | 2.04^                        |
| Not a parent            | 8.5%                           |                                         | 74.2%                          |                                         |                                     |                                              |                              |
| Parent                  |                                |                                         |                                |                                         |                                     |                                              |                              |
| Sample size             | 82                             | n.a.                                    | 31                             | n.a.                                    | n.a.                                | n.a.                                         | n.a.                         |

Notes: NP=near-peer. CE=community educator. n.a. = not applicable. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. Effect sizes of baseline differences > .05 are marked with ^.

## Appendix B – CONSORT diagrams

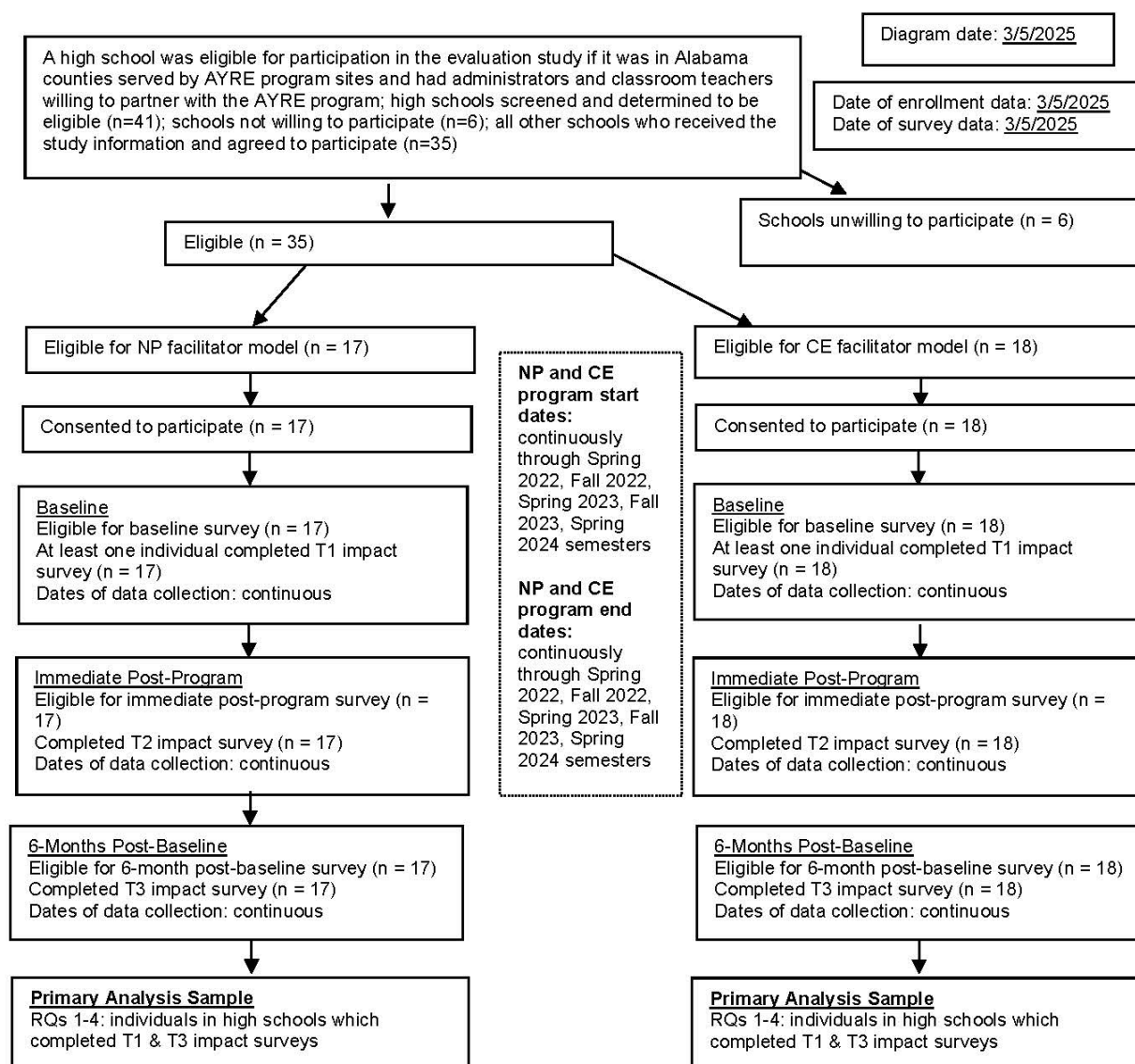
**Figure B.1. CONSORT diagram for individual clients, for studies in which consent occurred after assignment**



Notes: NP=near-peer. CE=community educator. T1=Time 1 survey (i.e., baseline). T2=T2 follow-up survey (i.e., immediate post-program). T3=Time 3 follow-up survey (i.e., 6-months post-baseline).

**Table B.6 Flow of individual participants through stages of the impact study, by condition**

| Stage of impact study               | NP facilitator model                             | CE facilitator model                             |
|-------------------------------------|--------------------------------------------------|--------------------------------------------------|
| Eligible for impact study           | 3,091                                            | 3,127                                            |
| Consented to participate            | 1,787                                            | 2,483                                            |
| <i>Eligible but didn't consent</i>  | <i>1,304</i>                                     | <i>644</i>                                       |
| T1 Survey                           |                                                  |                                                  |
| Eligible for T1                     | 1,787                                            | 2,483                                            |
| Completed T1                        | 1,455                                            | 2,067                                            |
| Dates of data collection            | Continuous                                       | Continuous                                       |
| <i>No response</i>                  | <i>332</i>                                       | <i>416</i>                                       |
| T2 Survey                           |                                                  |                                                  |
| Eligible for T2                     | 1,787                                            | 2,483                                            |
| Completed T2                        | 1,353                                            | 1,846                                            |
| Dates of data collection            | Continuous                                       | Continuous                                       |
| <i>No response</i>                  | <i>434</i>                                       | <i>637</i>                                       |
| T3 Survey                           |                                                  |                                                  |
| Eligible for T3                     | 1,787                                            | 2,483                                            |
| Completed T3                        | 1,180                                            | 1,559                                            |
| Dates of data collection            | Continuous                                       | Continuous                                       |
| <i>No response</i>                  | <i>607</i>                                       | <i>924</i>                                       |
| Primary Analysis Sample for RQs 1-4 | Individuals who completed T1 & T3 impact surveys | Individuals who completed T1 & T3 impact surveys |

**Figure B.2. CONSORT diagram for high schools (i.e., cluster-level), for studies in which consent occurred after assignment**

Notes: NP=near-peer. CE=community educator. T1=Time 1 survey (i.e., baseline). T2=T2 follow-up survey (i.e., immediate post-program). T3=Time 3 follow-up survey (i.e., 6-months post-baseline).

**Table B.7 Flow of high schools through stages of the impact study, by condition**

| Stage of impact study                | NP facilitator model                                          | CE facilitator model                                          |
|--------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|
| Eligible for impact study            | 17                                                            | 18                                                            |
| Consented to participate             | 17                                                            | 18                                                            |
| T1 Survey                            |                                                               |                                                               |
| Eligible for T1                      | 17                                                            | 18                                                            |
| At least one individual completed T1 | 17                                                            | 18                                                            |
| Dates of data collection             | Continuous                                                    | Continuous                                                    |
| T2 Survey                            |                                                               |                                                               |
| Eligible for T2                      | 17                                                            | 18                                                            |
| At least one individual completed T2 | 17                                                            | 18                                                            |
| Dates of data collection             | Continuous                                                    | Continuous                                                    |
| T3 Survey                            |                                                               |                                                               |
| Eligible for T3                      | 17                                                            | 18                                                            |
| At least one individual completed T3 | 17                                                            | 18                                                            |
| Dates of data collection             | Continuous                                                    | Continuous                                                    |
| Primary Analysis Sample for RQs 1-4  | Individuals in schools which completed T1 & T3 impact surveys | Individuals in schools which completed T1 & T3 impact surveys |

## Appendix C: Baseline Equivalence

### C.1. Baseline equivalence assessment

**Table C.1a Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals included in the analyses for RQ1 (original sample)**

| Baseline measure                                                          | NP facilitator model mean | NP facilitator model standard deviation | CE facilitator model mean | CE facilitator model standard deviation | NP and CE difference in means | p-value of test of difference in means | Effect size (absolute value) |
|---------------------------------------------------------------------------|---------------------------|-----------------------------------------|---------------------------|-----------------------------------------|-------------------------------|----------------------------------------|------------------------------|
| Sex                                                                       | 54.1%                     | n.a.                                    | 55.7%                     | n.a.                                    | 1.6%                          | .51                                    | .05                          |
| Female                                                                    | 45.9%                     |                                         | 44.3%                     |                                         |                               |                                        |                              |
| Male                                                                      |                           |                                         |                           |                                         |                               |                                        |                              |
| Race                                                                      |                           | n.a.                                    |                           | n.a.                                    | 37.7%                         | <.001                                  | .94 <sup>^</sup>             |
| Other race                                                                | 66.3%                     |                                         | 28.6%                     |                                         |                               |                                        |                              |
| White                                                                     | 33.7%                     |                                         | 71.4%                     |                                         |                               |                                        |                              |
| Relationship Status                                                       | 34.3%                     | n.a.                                    | 37.3%                     | n.a.                                    | 3%                            | .20                                    | .08 <sup>^</sup>             |
| Dating                                                                    | 65.7%                     |                                         | 62.7%                     |                                         |                               |                                        |                              |
| Not in a relationship                                                     |                           |                                         |                           |                                         |                               |                                        |                              |
| RQ1 Outcome: Relationship Skills Knowledge ( <i>value range 0 to 10</i> ) | 5.91                      | 1.67                                    | 6.11                      | 1.59                                    | .19                           | .02                                    | .12 <sup>^</sup>             |
| Sample size                                                               | 688                       | n.a.                                    | 996                       | n.a.                                    | n.a.                          | n.a.                                   | n.a.                         |

Notes: NP=near-peer. CE=community educator. n.a. = not applicable. See Table IV.3 for description of demographic covariates. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. We computed baseline equivalence tests with non-imputed data. Effect sizes of baseline differences > .05 are marked with <sup>^</sup>.

**Table C.1b Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals included in the analyses for RQ1 (matched sample, achieved via propensity score matching)**

| Baseline measure                                                          | NP facilitator model mean | NP facilitator model standard deviation | CE facilitator model mean | CE facilitator model standard deviation | NP and CE difference in means | p-value of test of difference in means | Effect size (absolute value) |
|---------------------------------------------------------------------------|---------------------------|-----------------------------------------|---------------------------|-----------------------------------------|-------------------------------|----------------------------------------|------------------------------|
| Sex                                                                       | 54.4%                     | n.a.                                    | 54.3%                     | n.a.                                    | .1%                           | 1                                      | .00                          |
| Female                                                                    | 45.6%                     |                                         | 45.7%                     |                                         |                               |                                        |                              |
| Male                                                                      |                           |                                         |                           |                                         |                               |                                        |                              |
| Race                                                                      | 55.1%                     | n.a.                                    | 55.1%                     | n.a.                                    | 0%                            | 1                                      | .00                          |
| Other race                                                                | 44.9%                     |                                         | 44.9%                     |                                         |                               |                                        |                              |
| White                                                                     |                           |                                         |                           |                                         |                               |                                        |                              |
| Relationship Status                                                       | 31.5%                     | n.a.                                    | 36.4%                     | n.a.                                    | 4.9%                          | .12                                    | .11^                         |
| Dating                                                                    | 68.5%                     |                                         | 63.6%                     |                                         |                               |                                        |                              |
| Not in a relationship                                                     |                           |                                         |                           |                                         |                               |                                        |                              |
| RQ1 Outcome: Relationship Skills Knowledge ( <i>value range 0 to 10</i> ) | 6.03                      | 1.65                                    | 5.88                      | 1.65                                    | .15                           | .15                                    | 0.09^                        |
| Sample size                                                               | 517                       | n.a.                                    | 517                       | n.a.                                    | n.a.                          | n.a.                                   | n.a.                         |

Notes: NP=near-peer. CE=community educator. n.a. = not applicable. See Table IV.3 for description of demographic covariates. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. We computed baseline equivalence tests with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

**Table C.1c Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals included in the analyses for RQ2 (original sample)**

| Baseline measure                                                       | NP facilitator model mean | NP facilitator model standard deviation | CE facilitator model mean | CE facilitator model standard deviation | NP and CE difference in means | p-value of test of difference in means | Effect size (absolute value) |
|------------------------------------------------------------------------|---------------------------|-----------------------------------------|---------------------------|-----------------------------------------|-------------------------------|----------------------------------------|------------------------------|
| Sex                                                                    | 56.6%                     | n.a.                                    | 56.1%                     | n.a.                                    | 0.5%                          | .81                                    | .02                          |
| Female                                                                 | 43.4%                     |                                         | 43.9%                     |                                         |                               |                                        |                              |
| Male                                                                   |                           |                                         |                           |                                         |                               |                                        |                              |
| Race                                                                   | 67.2%                     | n.a.                                    | 32.3%                     | n.a.                                    | 34.9%                         | <.001                                  | .91^                         |
| Other race                                                             | 32.8%                     |                                         | 67.7%                     |                                         |                               |                                        |                              |
| White                                                                  |                           |                                         |                           |                                         |                               |                                        |                              |
| Relationship Status                                                    | 33.1%                     | n.a.                                    | 36.4%                     | n.a.                                    | 3.3%                          | .11                                    | .11^                         |
| Dating                                                                 | 66.9%                     |                                         | 63.6%                     |                                         |                               |                                        |                              |
| Not in a relationship                                                  |                           |                                         |                           |                                         |                               |                                        |                              |
| RQ2 Outcome: Dating Aggression Tolerance ( <i>value range 1 to 4</i> ) | 1.54                      | .54                                     | 1.52                      | .52                                     | .03                           | .28                                    | .05                          |
| Sample size                                                            | 892                       | n.a.                                    | 1253                      | n.a.                                    | n.a.                          | n.a.                                   | n.a.                         |

Notes: NP=near-peer. CE=community educator. n.a. = not applicable. See Table IV.3 for description of demographic covariates. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. We computed baseline equivalence tests with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.



**Table C.1d Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals included in the analyses for RQ2 (matched sample, achieved via propensity score matching)**

| Baseline measure                                                       | NP facilitator model mean | NP facilitator model standard deviation | CE facilitator model mean | CE facilitator model standard deviation | NP and CE difference in means | p-value of test of difference in means | Effect size (absolute value) |
|------------------------------------------------------------------------|---------------------------|-----------------------------------------|---------------------------|-----------------------------------------|-------------------------------|----------------------------------------|------------------------------|
| Sex                                                                    | 57.4%                     | n.a.                                    | 54.5%                     | n.a.                                    | 2.9%                          | .32                                    | .07 <sup>^</sup>             |
| Female                                                                 | 42.6%                     |                                         | 45.5%                     |                                         |                               |                                        |                              |
| Male                                                                   |                           |                                         |                           |                                         |                               |                                        |                              |
| Race                                                                   | 58.1%                     | n.a.                                    | 58.1%                     | n.a.                                    | 0%                            | 1                                      | .00                          |
| Other race                                                             | 41.9%                     |                                         | 41.9%                     |                                         |                               |                                        |                              |
| White                                                                  |                           |                                         |                           |                                         |                               |                                        |                              |
| Relationship Status                                                    |                           | n.a.                                    | 34.1%                     | n.a.                                    | 2.8%                          | .29                                    | .08 <sup>^</sup>             |
| Dating                                                                 | 31.3%                     |                                         | 65.9%                     |                                         |                               |                                        |                              |
| Not in a relationship                                                  | 68.7%                     |                                         |                           |                                         |                               |                                        |                              |
| RQ2 Outcome: Dating Aggression Tolerance ( <i>value range 1 to 4</i> ) | 1.51                      | 0.53                                    | 1.59                      | 0.55                                    | .07                           | .01                                    | .13 <sup>^</sup>             |
| Sample size                                                            | 697                       | n.a.                                    | 697                       | n.a.                                    | n.a.                          | n.a.                                   | n.a.                         |

Notes: NP=near-peer. CE=community educator. n.a. = not applicable. See Table IV.3 for description of demographic covariates. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. We computed baseline equivalence tests with non-imputed data. Effect sizes of baseline differences > .05 are marked with <sup>^</sup>.

**Table C.1e Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals included in the analyses for RQ3 (original sample)**

| Baseline measure                                                      | NP facilitator model mean | NP facilitator model standard deviation | CE facilitator model mean | CE facilitator model standard deviation | NP and CE difference in means | p-value of test of difference in means | Effect size (absolute value) |
|-----------------------------------------------------------------------|---------------------------|-----------------------------------------|---------------------------|-----------------------------------------|-------------------------------|----------------------------------------|------------------------------|
| Sex                                                                   | 56.4%                     | n.a.                                    | 55.3%                     | n.a.                                    | 1.1%                          | .62                                    | .02                          |
| Female                                                                | 43.6%                     |                                         | 44.7%                     |                                         |                               |                                        |                              |
| Male                                                                  |                           |                                         |                           |                                         |                               |                                        |                              |
| Race                                                                  | 67.1%                     | n.a.                                    | 31.7%                     | n.a.                                    | 35.4%                         | <.001                                  | .91^                         |
| Other race                                                            | 32.9%                     |                                         | 68.3%                     |                                         |                               |                                        |                              |
| White                                                                 |                           |                                         |                           |                                         |                               |                                        |                              |
| Relationship Status                                                   | 36%                       | n.a.                                    | 37.5%                     | n.a.                                    | 1.5%                          | .48                                    | .08^                         |
| Dating                                                                | 64%                       |                                         | 62.5%                     |                                         |                               |                                        |                              |
| Not in a relationship                                                 |                           |                                         |                           |                                         |                               |                                        |                              |
| RQ3 Outcome: Relationship Self-Efficacy ( <i>value range 1 to 4</i> ) | 3.10                      | .5                                      | 3.10                      | .48                                     | .01                           | .62                                    | .02                          |
| Sample size                                                           | 892                       | n.a.                                    | 1265                      | n.a.                                    | n.a.                          | n.a.                                   | n.a.                         |

Notes: NP=near-peer. CE=community educator. n.a. = not applicable. See Table IV.3 for description of demographic covariates. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. We computed baseline equivalence tests with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

**Table C.1f Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals included in the analyses for RQ3 (matched sample, achieved via propensity score matching)**

| Baseline measure                                                      | NP facilitator model mean | NP facilitator model standard deviation | CE facilitator model mean | CE facilitator model standard deviation | NP and CE difference in means | p-value of test of difference in means | Effect size (absolute value) |
|-----------------------------------------------------------------------|---------------------------|-----------------------------------------|---------------------------|-----------------------------------------|-------------------------------|----------------------------------------|------------------------------|
| Sex                                                                   | 57.8%                     | n.a.                                    | 54.4%                     | n.a.                                    | 3.4%                          | .24                                    | .10 <sup>^</sup>             |
| Female                                                                | 42.2%                     |                                         | 45.6%                     |                                         |                               |                                        |                              |
| Male                                                                  |                           |                                         |                           |                                         |                               |                                        |                              |
| Race                                                                  | 57.7%                     | n.a.                                    | 57.7%                     | n.a.                                    | 0.0%                          | 1                                      | .00                          |
| Other race                                                            | 42.3%                     |                                         | 42.3%                     |                                         |                               |                                        |                              |
| White                                                                 |                           |                                         |                           |                                         |                               |                                        |                              |
| Relationship Status                                                   | 34.5%                     | n.a.                                    | 35.6%                     | n.a.                                    | 1.1%                          | .69                                    | .05                          |
| Dating                                                                | 65.5%                     |                                         | 64.4%                     |                                         |                               |                                        |                              |
| Not in a relationship                                                 |                           |                                         |                           |                                         |                               |                                        |                              |
| RQ3 Outcome: Relationship Self-Efficacy ( <i>value range 1 to 4</i> ) | 3.10                      | .51                                     | 3.09                      | .47                                     | .02                           | .54                                    | .03                          |
| Sample size                                                           | 693                       | n.a.                                    | 693                       | n.a.                                    | n.a.                          | n.a.                                   | n.a.                         |

Notes: NP=near-peer. CE=community educator. n.a. = not applicable. See Table IV.3 for description of demographic covariates. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. We computed baseline equivalence tests with non-imputed data. Effect sizes of baseline differences > .05 are marked with <sup>^</sup>.

**Table C.1g Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals included in the analyses for RQ4 (original sample)**

| Baseline measure                                     | NP facilitator model mean | NP facilitator model standard deviation | CE facilitator model mean | CE facilitator model standard deviation | NP and CE difference in means | p-value of test of difference in means | Effect size (absolute value) |
|------------------------------------------------------|---------------------------|-----------------------------------------|---------------------------|-----------------------------------------|-------------------------------|----------------------------------------|------------------------------|
| Sex                                                  | 56%                       | n.a.                                    | 55.6%                     | n.a.                                    | 0.4%                          | .86                                    | .00                          |
| Female                                               | 44%                       |                                         | 44.4%                     |                                         |                               |                                        |                              |
| Male                                                 |                           |                                         |                           |                                         |                               |                                        |                              |
| Race                                                 | 66.8%                     | n.a.                                    | 32.2%                     | n.a.                                    | 33.6%                         | <.001                                  | .89^                         |
| Other race                                           | 33.2%                     |                                         | 67.8%                     |                                         |                               |                                        |                              |
| White                                                |                           |                                         |                           |                                         |                               |                                        |                              |
| Relationship Status                                  | 33.4%                     | n.a.                                    | 36.6%                     | n.a.                                    | 3.2%                          | .17                                    | .11^                         |
| Dating                                               | 66.6%                     |                                         | 63.4%                     |                                         |                               |                                        |                              |
| Not in a relationship                                |                           |                                         |                           |                                         |                               |                                        |                              |
| RQ4 Outcome: Self-Regulation<br>(value range 1 to 5) | 3.60                      | .57                                     | 3.56                      | .56                                     | .05                           | .06                                    | .08^                         |
| Sample size                                          | 918                       | n.a.                                    | 1264                      | n.a.                                    | n.a.                          | n.a.                                   | n.a.                         |

Notes: NP=near-peer. CE=community educator. n.a. = not applicable. See Table IV.3 for description of demographic covariates. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. We computed baseline equivalence tests with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

**Table C.1h Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals included in the analyses for RQ4 (matched sample, achieved via propensity score matching)**

| Baseline measure                                     | NP facilitator model mean | NP facilitator model standard deviation | CE facilitator model mean | CE facilitator model standard deviation | NP and CE difference in means | p-value of test of difference in means | Effect size (absolute value) |
|------------------------------------------------------|---------------------------|-----------------------------------------|---------------------------|-----------------------------------------|-------------------------------|----------------------------------------|------------------------------|
| Sex                                                  | 57.6%                     | n.a.                                    | 54.0%                     | n.a.                                    | 3.6%                          | .21                                    | .10 <sup>^</sup>             |
| Female                                               | 42.4%                     |                                         | 46.0%                     |                                         |                               |                                        |                              |
| Male                                                 |                           |                                         |                           |                                         |                               |                                        |                              |
| Race                                                 | 57.2%                     | n.a.                                    | 57.2%                     | n.a.                                    | 0.0%                          | 1                                      | .00                          |
| Other race                                           | 42.8%                     |                                         | 42.8%                     |                                         |                               |                                        |                              |
| White                                                |                           |                                         |                           |                                         |                               |                                        |                              |
| Relationship Status                                  | 31.7%                     | n.a.                                    | 35.5%                     | n.a.                                    | 3.8%                          | .15                                    | .08 <sup>^</sup>             |
| Dating                                               | 68.3%                     |                                         | 64.5%                     |                                         |                               |                                        |                              |
| Not in a relationship                                |                           |                                         |                           |                                         |                               |                                        |                              |
| RQ4 Outcome: Self-Regulation<br>(value range 1 to 5) | 3.62                      | .57                                     | 3.54                      | .57                                     | .08                           | .01                                    | .14 <sup>^</sup>             |
| Sample size                                          | 710                       | n.a.                                    | 710                       | n.a.                                    | n.a.                          | n.a.                                   | n.a.                         |

Notes: NP=near-peer. CE=community educator. n.a. = not applicable. See Table IV.3 for description of demographic covariates. For tests of differences, we used chi-square tests for binary variables and *t*-tests for continuous variables. For effect sizes, we used Cox's index for binary variables and Hedge's *g* for continuous variables. We computed baseline equivalence tests with non-imputed data. Effect sizes of baseline differences > .05 are marked with <sup>^</sup>.

## C.2. Statistical approach to constructing equivalent groups

**Propensity Score Matching.** Since this study was a QED and baseline equivalence between groups in the raw sample was not achieved, we used propensity score matching as an approach to constructing equivalent groups. Propensity scores estimate the likelihood of belonging to each group and minimize the bias of non-random assignment (Caliendo & Kopeinig, 2008), and a matched sample can be obtained based on propensity scores (Austin, 2011). Propensity score matching is commonly used in QED studies and is widely accepted as an accurate method of testing program impact, superior to other methods such as individually controlling for key variables as covariates (Graham & Kurlaender, 2011). Specifically, we first estimated propensity scores using logistic regression, and then we used nearest-neighbor matching with caliper adjustment (0.2 *SD* of the logit of the propensity score; Austin, 2011; Harris & Horst, 2019) in the MatchIt package in R (Ho et al., 2011) to obtain a one-to-one matched sample of participants in the NP facilitator model and CE facilitator model, using absolute standardized difference to assess balance. The final propensity score matched sample for each primary RQ was obtained using race as the only covariate.

### C.3. Attrition Rates

**Table C.3. Overall and differential attrition rates at individual- and cluster-levels**

| Time point | Individual-Level Attrition     |                                |                                 |                   | Cluster-Level Attrition        |                                |                                 |                   |
|------------|--------------------------------|--------------------------------|---------------------------------|-------------------|--------------------------------|--------------------------------|---------------------------------|-------------------|
|            | NP facilitator model condition | CE facilitator model condition | Differential Between conditions | Overall Attrition | NP facilitator model condition | CE facilitator model condition | Differential Between conditions | Overall Attrition |
| T1         | 18.6%                          | 16.8%                          | 1.8%                            | 17.5%             | 0.0%                           | 0.0%                           | 0.0%                            | 0.0%              |
| T2         | 24.3%                          | 25.7%                          | 1.4%                            | 25.1%             | 0.0%                           | 0.0%                           | 0.0%                            | 0.0%              |
| T3         | 34.0%                          | 37.2%                          | 3.2%                            | 35.9%             | 0.0%                           | 0.0%                           | 0.0%                            | 0.0%              |

Notes: For number of surveys at each time point, see Table IV.1. We calculated individual-level attrition rates as the number of individuals without survey data / number of total individuals consented. We calculated differential attrition between conditions as the absolute value of the difference between NP facilitator model and CE facilitator model attrition rates. We calculated cluster-level attrition rates as the number of clusters (i.e., high schools) in which no individuals had survey data / number of total clusters in full sample. According to WWC Standards (2022), the optimistic boundary for differential attrition between conditions of a study with overall attrition of 36% (i.e., this study's overall attrition at the individual-level) is 7.0%; all individual-level differential attrition rates are below that cautious boundary. According to WWC Standards (2022), the optimistic boundary for differential attrition between conditions of a study with overall attrition of 0% (i.e., this study's overall attrition at the cluster-level) is 10.0%; all couple-level differential attrition rates are below that cautious boundary. Note that the WWC Standards (2022) optimistic boundary was used for this study since participants were high school students receiving programming in schools (per the HMRE Program Evaluation Toolkit, Friend et al., 2024). Thus, overall, AYRE classifies as a low attrition study at both the individual- and cluster-levels.

## Appendix D: Data Preparation

### D.1. Data Preparation Detailed

**Proactive measures during data collection.** During the startup period of this grant cycle, the local evaluation team piloted the AYRE impact surveys (multiple times and with a wide array of respondents & possible participant scenarios). We accordingly edited AYRE impact Qualtrics surveys for readability and for accuracy in data downloads (e.g., item names, values, etc.). During each workshop series, the evaluation team worked in real-time to troubleshoot any challenges participants experienced during the nFORM and AYRE impact survey data collection process. During each survey window period (i.e., baseline data collection days in high school classrooms, immediate post-program data collection days in high school classrooms, and T3 impact survey window for participants to complete on their own time), local evaluation staff were available to respond to program staff or participants' questions/challenges as they occurred. Throughout the impact study period, we regularly downloaded, merged, and checked AYRE impact survey and nFORM survey data for accuracy, outliers, and descriptive statistics. Immediate efforts were made to validate and/or correct data entries (e.g., if youth accidentally enter an incorrect ID), as well as to anticipate and address any other factors that may hinder accurate data collection.

**Data cleaning (after data collection was complete).** Completed T1 impact survey data for each academic semester (i.e., fall, spring) and T3 impact survey data for each monthly cohort were downloaded from Qualtrics into separate SPSS databases. nFORM Entrance survey data for each academic semester was also downloaded and imported into SPSS databases. For this impact evaluation, all relevant study data (i.e., T1 impact surveys, nFORM Entrance surveys, and T3 impact surveys) were merged based on participant individual ID in order to create a master dataset with survey responses from each of the monthly cohorts across each of the two timepoints (i.e., baseline and 6-months post-baseline). Basic descriptive statistics were run on all variables of interest to assess for outliers or impossible scores on each scale. Data plots (i.e., bar charts, scatter plot graphs) were used to look for outliers or improbable values. We also spot-checked data by randomly selecting cases and checking their raw scores against the overall dataset. After all error-checking, impossible scores were removed, and outliers were assessed for accuracy on a case-by-case basis. All adjustments for data errors were documented. Consistency in demographic data was assessed across participants and time points. Evaluation staff used a master database containing participant IDs and demographic data (obtained from both the nFORM Applicant Characteristics form and the T1 impact survey) to verify accuracy in demographic data across participants and timepoints. Responses that were obvious keystroke errors based on master database information and inconsistency with other timepoints were corrected. To ensure that all study outcome measures were performing as expected, we assessed the reliability of each measure (via coefficient alpha, as appropriate) and the associations among the measures were computed for each wave (via Pearson correlation coefficients).

**Missing Data Imputation.** Missing data in this study only occurred in analytic samples for RQ 3 and RQ4, since for RQ 1 and RQ 2 we kept only cases with complete data on all outcome items at baseline and 6-month follow-up. In the analytic samples for RQ 3 and RQ 4, missing data was present for cases missing 1-20% of items on relevant outcome measures at baseline or 6-month follow-up. Missing data were addressed through multiple imputation using the 'mice' package in RStudio (van Buuren & Groothuis-Oudshoorn, 2011; R Core Team, 2021). The 'mice' method ("Multivariate Imputation by Chained Equations") fills in missing values by using other observed data points as covariates, generating multiple plausible estimates based on the responses of all participants. To impute missing values, the classification and regression trees (CART) method was applied to the available raw data. The imputation process incorporated various covariates, including sex, race, age, relationship status, family structure, GPA, and other demographic factors. Ultimately, multiple imputation generated 10 imputed datasets, and the mean values from these datasets were aggregated into a single dataset for analysis.



**Table D.1. Internal consistency of outcome measures across timepoints**

| Outcome Measure             | T1  | T3  |
|-----------------------------|-----|-----|
| Dating Aggression Tolerance | .72 | .83 |
| Relationship Self-Efficacy  | .78 | .82 |
| Self-Regulation             | .74 | .81 |

Note: T1=Time 1 survey (i.e., baseline). T3=Time 3 follow-up survey (i.e., 6-months post-baseline). Internal consistency was calculated using Cronbach's alpha of all items in each measure. Relationship Skills Knowledge outcome was a quiz; a sum index composite score of correct items (out of 10) was calculated, so scale internal consistency was not viable.

**Table D.2. Correlations Among Outcome Measures Across Timepoints**

|                                     | 1       | 2       | 3       | 4       | 5      | 6      | 7      | 8    |
|-------------------------------------|---------|---------|---------|---------|--------|--------|--------|------|
| 1. T1 Relationship Skills Knowledge | 1       |         |         |         |        |        |        |      |
| 2. T3 Relationship Skills Knowledge | .44***  | 1       |         |         |        |        |        |      |
| 3. T1 Dating Aggression Tolerance   | -.31*** | -.25*** | 1       |         |        |        |        |      |
| 4. T3 Dating Aggression Tolerance   | -.21*** | -.32*** | .33***  | 1       |        |        |        |      |
| 5. T1 Relationship Self-Efficacy    | .06**   | -.01    | -.12*** | -.12*** | 1      |        |        |      |
| 6. T3 Relationship Self-Efficacy    | .12***  | .16***  | -.13*** | -.19*** | .35*** | 1      |        |      |
| 7. T1 Self-Regulation               | .24***  | .11***  | -.24*** | -.16*** | .27*** | .21*** | 1      |      |
| 8. T3 Self-Regulation               | .22***  | .31***  | -.20*** | -.28*** | .19*** | .34*** | .42*** | 1    |
| Mean                                | 5.87    | 6.19    | 1.60    | 1.46    | 3.10   | 3.10   | 3.55   | 3.63 |
| SD                                  | 1.66    | 1.96    | .58     | .56     | .49    | .50    | .60    | .60  |

Notes: T1=Time 1 survey (i.e., baseline). T3=Time 3 follow-up survey (i.e., 6-months post-baseline). SD=standard deviation. Relationship Skills Knowledge outcome (sum composite; higher scores represent greater knowledge of healthy relationship skills); Dating Aggression Tolerance outcome (mean composite; higher scores represent higher tolerance of dating aggression). Relationship Self-Efficacy outcome (mean composite; higher scores represent more self-efficacy in relationships). Self-Regulation outcome (mean composite; higher scores represent better self-regulation). Pearson correlation coefficients were utilized to check for associations between outcome measures across time points. \*\* $p < .01$ . \*\*\* $p < .001$ .

**Table D.3. Correlations between covariates and outcome measures**

|                        | T1<br>Relationship<br>Skills<br>Knowledge | T3<br>Relationship<br>Skills<br>Knowledge | T1 Dating<br>Aggression<br>Tolerance | T3 Dating<br>Aggression<br>Tolerance | T1<br>Relationship<br>Self-Efficacy | T3<br>Relationship<br>Self-Efficacy | T1 Self-<br>Regulation | T3 Self-<br>Regulation |
|------------------------|-------------------------------------------|-------------------------------------------|--------------------------------------|--------------------------------------|-------------------------------------|-------------------------------------|------------------------|------------------------|
| Sex                    | -.15***                                   | -.12***                                   | .24***                               | .15***                               | -.01                                | -.05*                               | -.05**                 | -.10***                |
| Race                   | -.22***                                   | -.21***                                   | .16***                               | .09***                               | -.02                                | -.06**                              | -.01                   | -.03                   |
| Relationship<br>Status | -.08***                                   | -.06**                                    | .05**                                | -.01                                 | .15***                              | .09***                              | .03                    | -.02                   |

Notes: T1=Time 1 survey (i.e., baseline). T3=Time 3 follow-up survey (i.e., 6-months post-baseline). Relationship Skills Knowledge outcome (sum composite; higher scores represent greater knowledge of healthy relationship skills); Dating Aggression Tolerance outcome (mean composite; higher scores represent higher tolerance of dating aggression). Relationship Self-Efficacy outcome (mean composite; higher scores represent more self-efficacy in relationships). Self-Regulation outcome (mean composite; higher scores represent better self-regulation). Pearson correlation coefficients were utilized to check for associations between outcome measures and covariates. See Table IV.3 for description of demographic covariates. \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

**Table D.4. Missing data on each outcome for each RQ analytic sample, separately by condition**

|    |                               | NP facilitator model              |                                                                                  |                 |                 |                  | CE facilitator model              |                                                                                  |                 |                 |                  |
|----|-------------------------------|-----------------------------------|----------------------------------------------------------------------------------|-----------------|-----------------|------------------|-----------------------------------|----------------------------------------------------------------------------------|-----------------|-----------------|------------------|
| RQ | Outcome Measure               | # Total cases in full HMRF sample | # Cases retained for analytic sample (missing <=20% of outcome items on T1 & T3) | # Cases dropped | # Cases imputed | Missingness Rate | # Total cases in full HMRF sample | # Cases retained for analytic sample (missing <=20% of outcome items on T1 & T3) | # Cases dropped | # Cases imputed | Missingness Rate |
| 1  | Relationship Skills Knowledge | 969                               | 688                                                                              | 281             | 0               | 30.0%            | 1275                              | 996                                                                              | 279             | 0               | 21.9%            |
| 2  | Dating Aggression Tolerance   | 969                               | 892                                                                              | 77              | 0               | 7.9%             | 1275                              | 1253                                                                             | 22              | 0               | 1.7%             |
| 3  | Relationship Self-Efficacy    | 969                               | 892                                                                              | 77              | 40              | 7.9%             | 1275                              | 1265                                                                             | 10              | 42              | 0.8%             |
| 4  | Self-Regulation               | 969                               | 918                                                                              | 51              | 48              | 5.3%             | 1275                              | 1264                                                                             | 11              | 57              | 0.9%             |

Notes: T1=Time 1 survey (i.e., baseline). T3=Time 3 follow-up survey (i.e., 6-months post-baseline). For each condition, the # of total cases in full HMRF sample represent the # of cases with survey data on T1 impact, nFORM Entrance, and T3 impact surveys (see Table IV.2 for baseline equivalence assessment of this full HMRF analytic sample). Relationship Skills Knowledge (i.e., 10 total items; since this is an index sum composite of correct answers, cases missing 1+ items on either T1 or T3 impact surveys were dropped; since only cases with 100% of outcome items were retained, no cases with missing data were imputed). Dating Aggression Tolerance (i.e., 4 total items; cases missing 1+ items on either T1 or T3 impact surveys were dropped; since only cases with 100% of outcome items were retained, no cases with missing data were imputed). Relationship Self-Efficacy (i.e., 6 total items; cases missing 2+ items on either T1 or T3 impact surveys were dropped; of the cases retained for analytic sample, those missing 1 item on either T1 or T3 impact surveys were imputed). Self-Regulation (i.e., 10 total items; cases missing 3+ items on either T1 or T3 impact surveys were dropped; of the cases retained for analytic sample, those missing 1-2 items on either T1 or T3 impact surveys were imputed). Missingness rate is a percentage, calculated as # of cases dropped / # total cases in full HMRF sample. See Table IV.1 for overview of individual sample sizes by condition and outcome.

## Appendix E: Impact estimation

### E.1. Impact estimation steps detailed

This section describes the impact estimation methods used to analyze data for the study. Analyses were run in SPSS and R online statistical software.

**Preliminary Analyses.** Before applying the modeling approach outlined below, we verified model assumptions (e.g., normal distribution of primary measures, etc.), and found that no variable transformations were necessary. We examined Pearson correlation coefficients for all primary outcomes across time points, with all separate T1 and T3 outcome measures showing correlations above .30 (see Appendix D, Table D.2). Additionally, correlations between potential covariates and primary outcomes were assessed (see Table IV.3 for description of covariates; see Appendix D, Table D.3 for correlations between outcomes and covariates). Demographic covariate selection was based on their potential relevance to the primary outcomes.

**Modeling approach.** For each primary RQ, we used the following procedures in order:

1. **Preliminary Assessment of Baseline Equivalence:** Separately for each RQ's analytic sample, we assessed baseline equivalence on demographic characteristics and the relevant outcome according to WWC Standards (2022; see Chapter IV, Section IV.B above).
2. **Propensity Score Matching:** Since baseline differences between conditions in all four analytic samples had some effect sizes greater than 0.05, we used propensity score matching separately for each analytic sample to ensure adequate balance between the two study groups (according to the HMRE Program Evaluation Toolkit, Friend et al., 2024). Specifically, we first estimated propensity scores using logistic regression, and then we used nearest-neighbor matching with caliper adjustment (0.2 *SD* of the logit of the propensity score; Austin, 2011; Harris & Horst, 2019) in the MatchIt package in R (Ho et al., 2011) to obtain a one-to-one matched sample of participants in the NP facilitator model and CE facilitator model. We obtained the final propensity score matched sample for each primary RQ by using race as the only covariate (see Appendix C, Section C.2 for details about propensity score matching methods used).
3. **Secondary Assessment of Baseline Equivalence:** Separately for each RQ analytic sample, we re-assessed baseline equivalence on demographic characteristics and the relevant outcome, using the matched sample obtained in #2 above. For each of the four analytic samples, some moderate baseline differences (i.e., effect sizes between .05 and .25) in demographics and/or baseline levels of the relevant outcome remained in the matched samples (Tables C.1b, C.1d, C.1f, C.1h). According to WWC Standards (2022), moderate baseline differences are eligible for adjustment (see following Step #4).
4. **Adjusted Outcome Means (Using Matched Sample):** Since the secondary assessment of baseline equivalence on the matched sample did not show equivalence between groups on the relevant outcomes in analytic samples for RQs 1, 2, 3, and 4 (see #3 above), we used a linear regression model (including the demographic variables, baseline levels of the outcome, and condition [NP or CE] as independent variables, and the T3 outcome as dependent variable) to calculate adjusted predicted values of the relevant outcome.
5. **Equivalent Effects Testing (Using Matched Sample and Adjusted Outcome Means):** For RQs 1, 2, 3, and 4, we used the adjusted outcome means with the propensity score matched sample for each analytic sample to conduct the following equivalent effects testing procedures (outlined in the *HMRE Impact Analysis Plan Instructions Guide* and Lakens, 2017). In reviewing relevant research on the impact of youth relationship education on outcomes similar to those in this study (i.e., relationship knowledge, dating aggression tolerance, relationship self-efficacy, and

self-regulation), we decided on a pre-specified “smallest effect size of interest” (SESOI) of  $d = 0.20$ . With this SESOI, we defined the difference in effect size equivalence interval ( $\Delta_L$ ,  $\Delta_U$ ), where  $\Delta_L$  was the lower bound and  $\Delta_U$  is the upper bound of the difference in effect sizes between groups. For each Primary RQ (and relevant outcome), the alternative hypothesis was that the difference in effect sizes of the two groups is inside the equivalence interval (either  $> \Delta_L$  or  $< \Delta_U$ ). The null hypothesis was that the difference in effect sizes of the two groups is outside the equivalence interval (either  $\leq \Delta_L$  or  $\geq \Delta_U$ ). For each Primary RQ, we used Lakens et al. (2018) package in R online statistical software to calculate the difference in group means of outcomes at T3 impact survey with  $t$ -tests to conduct two one-sided significance tests (i.e., TOST procedure) with the relevant null and alternative hypotheses. Conclusions were made about whether the difference in the group means between the two conditions was statistically equivalent to zero or not. Findings were considered statistically significant based on  $p < .05$ , two-tailed test.

**Model specification.** Covariates were included in propensity score matching analyses if they exhibited baseline inequivalence between conditions (see Table IV.3 for description of covariates). Covariates, baseline levels of outcomes, and condition (i.e., NP facilitator model or CE facilitator model) were included in linear regression models to calculate adjusted outcome means. Outcome variables for the equivalent effects testing included relationship skills knowledge (RQ 1), dating aggression tolerance (RQ 2), relationship self-efficacy (RQ 3), and self-regulation (RQ 4). Refer to Appendix E for the model equations used to estimate impacts.

## E.2 Power Analyses for each RQ sample

Post-hoc power analyses were conducted in nQuery online software separately for each Primary RQ analytic sample to assess power available to detect effects using the TOST equivalent effects testing procedures (Julious, 2004; Lakens, 2017). Using a true mean difference in targeted outcomes between the two conditions of zero, the pooled standard deviation of each targeted outcome, an SESOI of  $d=0.20$ , and a significance level of  $\alpha=.05$ , the matched sample size for RQ 1 ( $n=517$  in each condition) had 88% power to detect effects; RQ 2 sample ( $n=697$  in each condition) had 97% power to detect effects; RQ 3 sample ( $n=693$  in each condition) had 95% power to detect effects; and RQ 4 sample ( $n=710$  in each condition) had 96% power to detect effects.

## E.3 Impact estimates by RQ detailed

**Relationship Skills Knowledge.** Results of the TOST equivalent effects test with the matched sample (obtained from propensity score matching) and adjusted outcome means (using the linear regression model as specified above in Chapter IV, Section IV.C), using a SESOI of 0.20 and equivalence interval based on pooled SD of (-0.180, 0.180), indicated that the two program conditions were not equivalent on relationship skills knowledge at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound = 0.381;  $p$  value test of difference in means greater or equal to upper bound  $<.001$ ; NP  $M=5.991$ ,  $SD=0.919$ ,  $n=517$ ; CE  $M=6.154$ ,  $SD=0.884$ ,  $n=517$ ; see Table V.1). Note that the results of equivalent effects testing are dependent on the adjusted outcome means calculated in the linear regression model as specified (see Chapter IV, Section IV.C above). We conducted a follow-up independent sample  $t$ -test, comparing the adjusted outcome means at 6-months post-baseline between NP and CE conditions; the CE youth on average reported statistically significantly higher relationship skills knowledge than NP youth ( $t(1032)=-2.80$ ,  $p=.005$ ).

**Dating Aggression Tolerance.** Results of the TOST equivalent effects test with the matched sample (obtained from propensity score matching) and adjusted outcome means (using the linear regression model as specified above in Chapter IV, Section IV.C), using an SESOI of 0.20 and

equivalence interval based on pooled SD of (-0.037, 0.037), indicated that the two program conditions were equivalent on dating aggression tolerance at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound  $<.001$ ;  $p$  value test of difference in means greater or equal to upper bound  $<.001$ ; NP  $M=1.460$ ,  $SD=0.177$ ,  $n=697$ ; CE  $M=1.457$ ,  $SD=0.189$ ,  $n=697$ ; see Table V.1). Note that the results of equivalent effects testing are dependent on the adjusted outcome means calculated in the linear regression model as specified (see Chapter IV, Section IV.C above).

**Relationship Self-Efficacy.** Results of the TOST equivalence tests with the matched sample (obtained from propensity score matching) and adjusted outcome means (using the linear regression model as specified above in Chapter IV, Section IV.C), using an SESOI of 0.20 and equivalence interval based on pooled SD of (-0.034, 0.034), indicated that the two program conditions were not equivalent on relationship self-efficacy at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound  $<.001$ ;  $p$  value test of difference in means greater or equal to upper bound  $=.149$ ; NP  $M=3.125$ ,  $SD=0.178$ ,  $n=693$ ; CE  $M=3.101$ ,  $SD=0.164$ ,  $n=693$ ; see Table V.1). Note that the results of equivalent effects testing are dependent on the adjusted outcome means calculated in the linear regression model as specified (see Chapter IV, Section IV.C above). We conducted a follow-up independent sample  $t$ -test, comparing the adjusted outcome means at 6-months post-baseline between NP and CE conditions; the CE youth on average reported statistically significantly higher relationship self-efficacy than NP youth ( $t(1304)=2.57$ ,  $p=.01$ ).

**Self-Regulation.** Results of the TOST equivalent effects test with the matched sample (obtained from propensity score matching) and adjusted outcome means (using the linear regression model as specified above in Chapter IV, Section IV.C), using an SESOI of 0.20 and equivalence interval based on pooled SD of (-0.047, 0.047), indicated that the two program conditions were equivalent on self-regulation at 6-months post-baseline ( $p$  value test of difference in means lower or equal to lower bound  $<.001$ ;  $p$  value test of difference in means greater or equal to upper bound  $<.001$ ; NP  $M=3.627$ ,  $SD=0.235$ ,  $n=710$ ; CE  $M=3.630$ ,  $SD=0.240$ ,  $n=710$ ; see Table V.1). Note that the results of equivalent effects testing are dependent on the adjusted outcome means calculated in the linear regression model as specified (see Chapter IV, Section IV.C above).

### E.3. Impact estimation equations

**Baseline Equivalence.** Hedge's  $g$  effect size was used to assess baseline equivalence on continuous demographics/outcomes, and Cox's index effect size was used to assess baseline equivalence on binary demographics.

The equation for **Hedge's  $g$  effect size** calculation is:

$$g = \frac{\overline{X}_1 - \overline{X}_2}{S^*}$$

where

- $\overline{X}_1$  = Mean of group 1
- $\overline{X}_2$  = Mean of group 2
- $S^*$  = Pooled standard deviation with correction for small sample bias

The pooled standard deviation is calculated as:

$$S^* = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

where

- $S_1^2$  = Variance of group 1
- $S_2^2$  = Variance of group 2
- $n_1$  = Sample size of group 1
- $n_2$  = Sample size of group 2

To correct for small sample bias, Hedge's  $g$  is adjusted using J correction:

$$g_{\text{corrected}} = g \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

where

- $n_1$  = Sample size of group 1
- $n_2$  = Sample size of group 2

The equation for **Cox's index effect size** calculation is:

$$g = [\log(\text{odds intervention}) - \log(\text{odds control})]/1.65$$



**Equivalent Effects Testing using Two One-Sided T-Tests (TOST).** The TOST procedure is an approach used in equivalence testing to determine if a treatment effect is statistically equivalent to a reference value within a defined equivalence margin. Below are the key components, including equations and variable descriptions (Microsoft, 2025):

## 1. Hypotheses for TOST Equivalence Testing

The goal of TOST is to test whether the true population mean ( $\mu$ ) lies within an acceptable range ( $\Delta$ ) around a reference value ( $\mu_0$ ).

**Null and Alternative Hypotheses:**

$$H_0 : (\mu - \mu_0 \leq -\Delta) \quad \text{or} \quad (\mu - \mu_0 \geq \Delta)$$

$$H_1 : -\Delta < \mu - \mu_0 < \Delta$$

where:

- $H_0$  = The effect is **not equivalent** (mean difference is outside the equivalence margin).
- $H_1$  = The effect is **equivalent** (mean difference is within the equivalence margin).

## 2. Test Statistics

The TOST procedure involves two one-sided t-tests:

### (a) Lower Bound Test

Tests if the mean is greater than the lower bound ( $-\Delta$ ):

$$t_{\text{lower}} = \frac{\bar{X} - \mu_0 + \Delta}{\frac{s}{\sqrt{n}}}$$

### (b) Upper Bound Test

Tests if the mean is less than the upper bound ( $+\Delta$ ):

$$t_{\text{upper}} = \frac{\bar{X} - \mu_0 - \Delta}{\frac{s}{\sqrt{n}}}$$

where:

- $t_{\text{lower}}$  = Test statistic for the lower bound test
- $t_{\text{upper}}$  = Test statistic for the upper bound test
- $\bar{X}$  = Sample mean
- $\mu_0$  = Population or reference mean
- $\Delta$  = Equivalence margin
- $s$  = Sample standard deviation
- $n$  = Sample size

### 3. Decision Rule

The null hypothesis of non-equivalence is rejected if **both tests are statistically significant** at the  $\alpha$  level:

$$t_{\text{lower}} > -t_{(n-1), 1-\alpha}$$

$$t_{\text{upper}} < t_{(n-1), 1-\alpha}$$

where:

- $t_{(n-1), 1-\alpha}$  = Critical value from the t-distribution with  $n - 1$  degrees of freedom at the significance level  $\alpha$