

SAMPLE DELIVERABLE

Local Evaluation Findings Report

Title V Sexual Risk Avoidance Education (SRAE) Program
Analysis of Participant Entry and Exit Surveys, Middle School Cohorts

16 cohorts | N = 367 surveys | Analysis-to-report turnaround: 5 business days

All client, site, curriculum, and location identifiers have been removed or renumbered. Every statistic is from a real delivered engagement.

Prepared by

[Your Organization's Name Here]

Every deliverable ships under your branding and your authorship. This line is where yours goes.

Sample by Intellectus 360 | Info@Intellectus360.com | (888) 3383-6639

Executive Summary

This report analyzes 367 anonymous participant surveys (196 entry, 171 exit) collected across 16 cohorts of a federally funded Sexual Risk Avoidance Education (SRAE) program delivered by a youth-serving nonprofit grantee in a high-need urban region. All cohorts completed the middle school version of the federal survey. Because three item families exist (entry-only, exit-only, and items asked at both time points), each family is described separately, and formal pre and post comparisons are restricted to the five life-skills knowledge items administered identically at entry and exit.

Key findings. First, participants entered the program with strong baseline knowledge; roughly 84 to 92 percent already answered the life-skills knowledge questions correctly at entry, leaving little room for measurable growth, and no entry-to-exit change on these items approached statistical significance. Second, the exit-only items tell a more encouraging story; between 62 and 83 percent of exiting participants reported that the program made them much or somewhat more likely to use core life skills, and 62 to 82 percent reported strengthened intentions to delay sexual activity and to sequence adult milestones (education, employment, marriage, children). Third, program experience ratings were strong; roughly three quarters of exiting youth reported feeling interested, respected, and able to ask questions all or most of the time. Fourth, a comparison of the three cohorts delivered in community settings (after-school and summer programs) with the thirteen school-based cohorts found no statistically significant differences on any measure, although the community cohorts trended slightly less favorable on every exit composite; their samples are very small, so this pattern is presented as a monitoring signal rather than a conclusion.

The discussion chapter addresses ceiling effects, the structural limits of the survey design, data-capture inconsistencies found and corrected during analysis, and concrete recommendations for strengthening the next evaluation cycle.

1. Background

The grantee delivers a Title V SRAE program using an evidence-based curriculum in schools, after-school programs, and community settings. The federal SRAE program is authorized under Section 510 of the Social Security Act (42 U.S.C. § 710), as amended by the Bipartisan Budget Act of 2018, and is implemented through a Positive Youth Development framework that emphasizes personal responsibility, self-regulation, goal setting, healthy relationships, and the avoidance of youth risk behaviors, with teen pregnancy prevention as the downstream outcome of interest (Family and Youth Services Bureau [FYSB], 2020). The framework's core constructs derive from the positive youth development literature (Catalano et al., 2004).

Each delivered group of youth constitutes a cohort. Participants complete an anonymous entry survey before programming and an anonymous exit survey afterward. This report covers 16 cohorts, renumbered 1 through 16 for de-identification, spanning school classrooms, after-school groups, and summer camps. Three cohorts (4, 11, and 16) served youth outside school classrooms and form the comparison group for the setting analysis.

2. Methods

2.1 Design and data

The evaluation uses a single-group entry/exit design with anonymous surveys, a common design for local program evaluation whose inferential limits are well described in the quasi-experimental literature (Shadish et al., 2002). Because surveys carry no participant identifier, entry and exit responses cannot be linked within person; all entry versus exit comparisons therefore treat the two waves as independent samples drawn from the same cohorts. Thirty cohort workbooks (15 entry, 15 exit) were combined into a single analytic file of 367 records. Two cohorts are unmatched in the available data: one contributed only entry surveys ($n = 14$) and one only exit surveys ($n = 9$). They are retained in descriptive analyses and drop out of any change computation automatically.

During data preparation, two source-file inconsistencies were identified and corrected. Eleven entry workbooks used a four-row header layout rather than the five-row layout of the remaining files; a per-file header detection routine was applied so that every record aligned to the correct question codes, and this recovered eleven participant records that a naive combination would have dropped. In addition, several items exist only on the high school survey; because every survey in this dataset is the middle school version, those columns are structurally empty and are excluded.

2.2 The three item families

Entry-only items measure baseline status: past-3-month substance use (yes/no), frequency of applying self-regulation skills (1 = all of the time to 4 = none of the time), and self-beliefs about goals and relationships (1 = not true at all to 3 = very true).

Exit-only items measure program-attributed change: whether being in the program made the participant more or less likely to use each skill or pursue each intention (1 = much more likely to 5 = much less likely), and program experience (1 = all of the time to 4 = none of the time). A point of method that matters: several exit items share question numbers with entry items but ask different questions on a different scale; they are retrospective attribution items, not re-measurements, and are analyzed as exit-only. Retrospective self-assessments of change carry known biases, including response-shift effects, and are interpreted accordingly (Howard, 1980). Response shift refers to the way an intervention can change the very yardstick a participant uses to rate themselves: a student who learns during the program what a healthy relationship actually involves may judge their earlier understanding more harshly, or their current understanding more generously, than they would have before the program. Because the internal standard moves between the two ratings, a retrospective answer to the question of how much one has changed is not equivalent to the difference between a true pretest and posttest, and can overstate or understate real change.

Items asked at both time points are five life-skills knowledge questions (internet safety, phone contacts, uncomfortable online contact, healthy friendships, and harmful messages), answered Yes, Unsure, or No, plus topic checkboxes and the demographic items. These are the only defensible basis for formal pre/post testing.

2.3 Scoring and statistical approach

Knowledge items were scored Yes = 1, Unsure = 0.5, No = 0, and a knowledge composite was computed as the mean of the five items for respondents answering at least three. Composites were also formed for entry skills frequency, entry self-beliefs, exit program-attributed change, exit future intentions, and exit program experience.

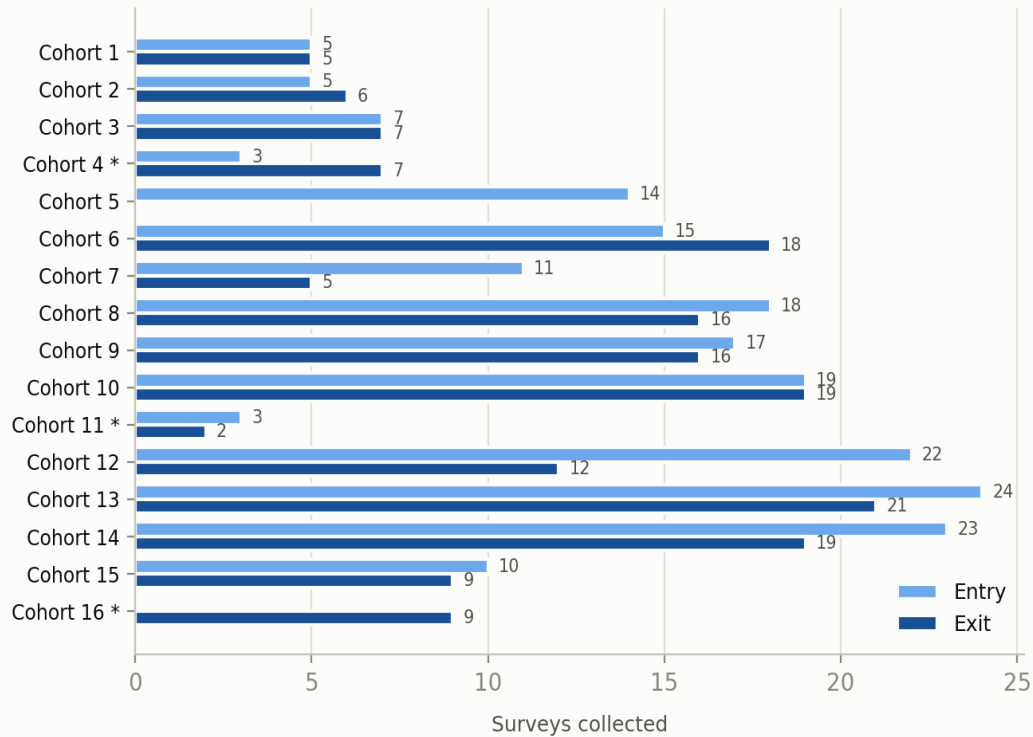
Entry versus exit proportions were tested with chi-square tests, or Fisher's exact test when any expected cell was small, with Benjamini-Hochberg false discovery rate adjustment across the knowledge items (Benjamini & Hochberg, 1995). Score and composite comparisons used Welch's t tests (Welch, 1947) with Cohen's d effect sizes (Cohen, 1988), with Mann-Whitney U as a sensitivity check (Mann & Whitney, 1947). For the setting comparison, the three community cohorts formed one group and the school-based cohorts the other; differential change on the knowledge composite was tested with a time by group interaction in a two-way linear model, within-group differences with Welch's t, and between-group differences at each time point with Welch's t. Alpha was 0.05, two-sided.

3. Results

3.1 Sample

Figure 1 shows the 367 surveys by cohort and wave. Cohort sizes range from 2 to 24 surveys per wave. The three community cohorts contributed 6 entry and 18 exit surveys, about 3 percent of the entry wave and 11 percent of the exit wave.

Figure 1. Surveys collected by cohort and time point



* Cohorts delivered in community settings (after-school and summer programs). N = 367 surveys. Cohorts renumbered for de-identification.

Table 1. Participant characteristics at entry and exit

Characteristic	Entry	Exit
Age, mean (SD)	11.74 (0.88), n=195	12.08 (1.03), n=170
Grade, mean (SD)	6.65 (0.7)	6.77 (0.84)
Female, %	57.7%	54.9%
Hispanic/Latino, %	72%	70.7%
Black or African American, %	21.3%	18.9%
White, %	11.5%	12.6%
Other race, %	59%	61.6%
English spoken, %	69.4%	73.8%
Spanish spoken, %	56.6%	57.7%
Living with family, %	95.3%	98.8%

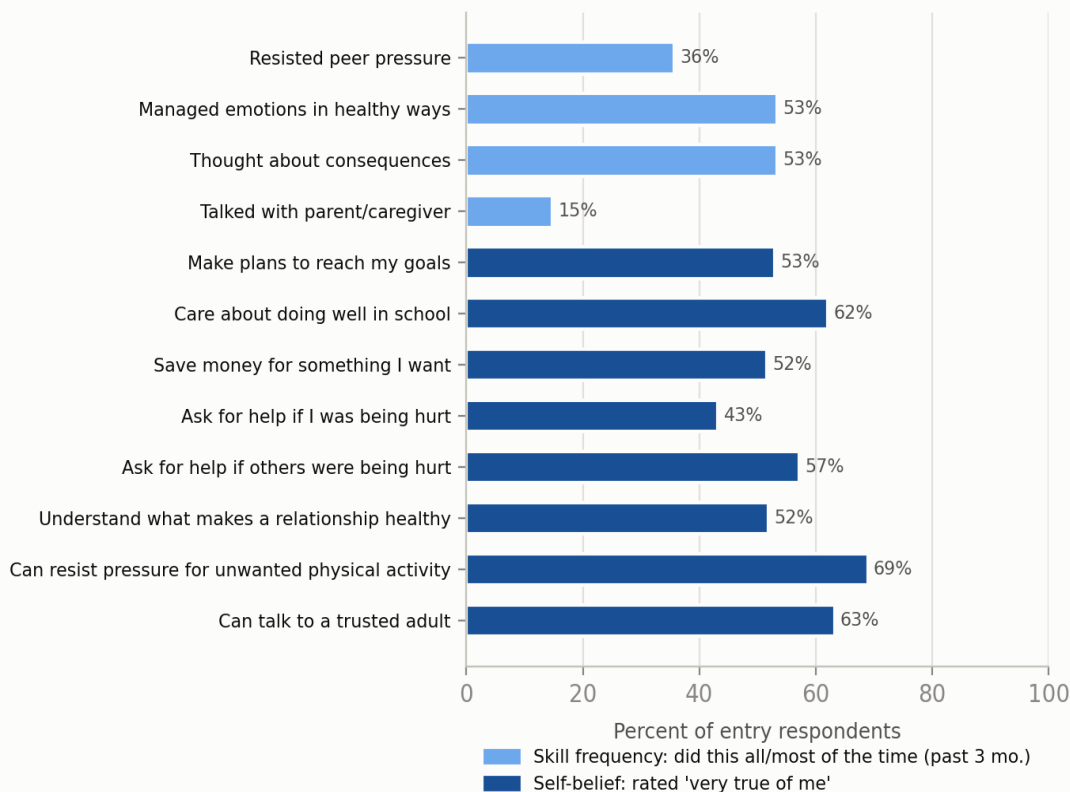
Race categories are multi-select and do not sum to 100. Entry and exit are unmatched cross-sections of the same cohorts, so small differences reflect sampling rather than change.

The demographic profiles of the two waves are closely similar, which supports treating them as draws from the same underlying population. Participants averaged roughly 12 years of age, about grade 7, predominantly Hispanic/Latino and/or Black, and nearly all lived with family.

3.2 Baseline: entry-only measures

Substance use at entry was low, consistent with a middle school population: 14 percent reported any past-3-month use of the six substances asked about, with vaping and alcohol the most common individual items. Self-regulation skills showed room for growth; only 36 percent said they resisted peer pressure all or most of the time in the prior three months, and 15 percent talked with a parent or caregiver about important decisions that often. Self-beliefs were moderate; roughly half to two thirds rated core goal and relationship statements as very true of themselves (Figure 2).

Figure 2. Baseline (entry-only) skills and assets

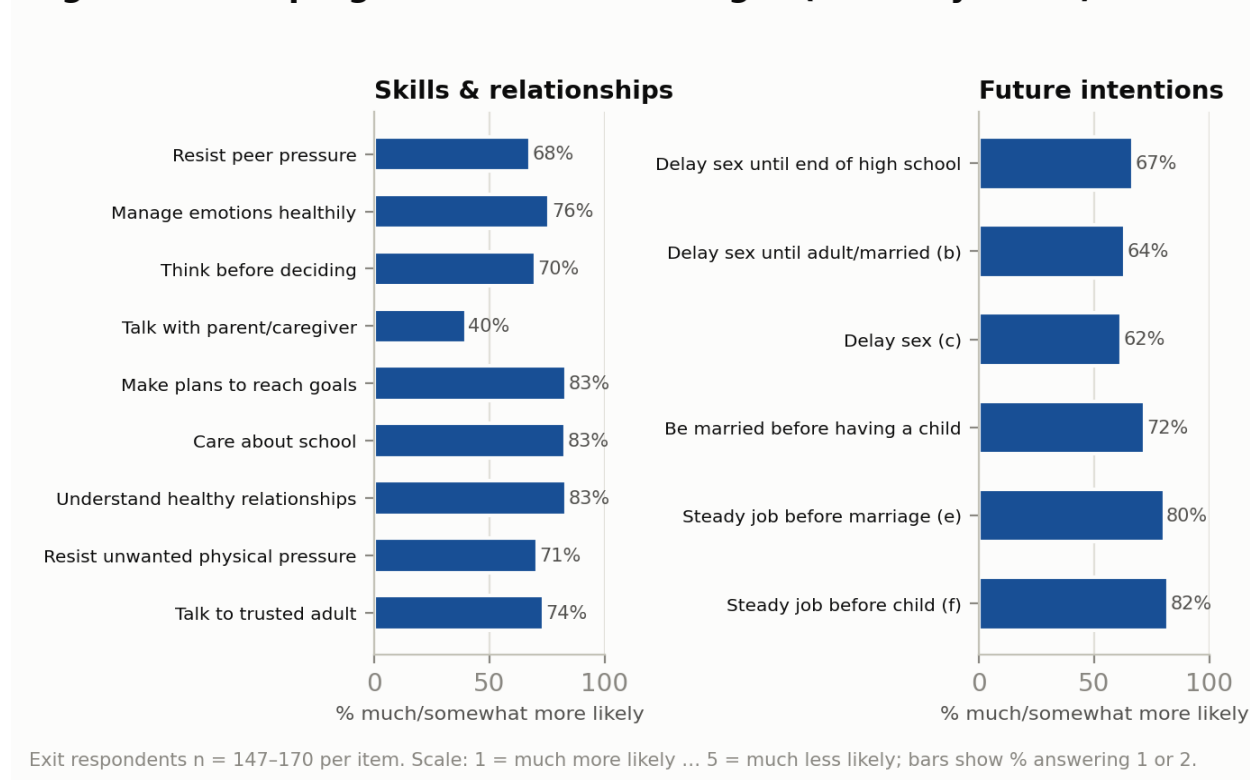


Entry n = 183-196 per item. Any past-3-month substance use at entry: 14% of respondents.

3.3 Exit-only measures: program-attributed change

On exit, majorities attributed positive change to the program across nearly every skill and intention (Figure 3). Between 68 and 83 percent said the program made them much or somewhat more likely to use core skills such as planning for goals, caring about school, and understanding healthy relationships. Intentions related to the program's central aims were also strong: 67 percent reported being more likely to delay sexual intercourse until at least the end of high school, and about 72 to 82 percent reported being more likely to complete milestones such as a steady job and marriage before having children. The lowest attribution item, at 40 percent, was talking with a parent or caregiver, mirroring the weak baseline for the same behavior.

Figure 3. Exit: program-attributed changes (exit-only items)

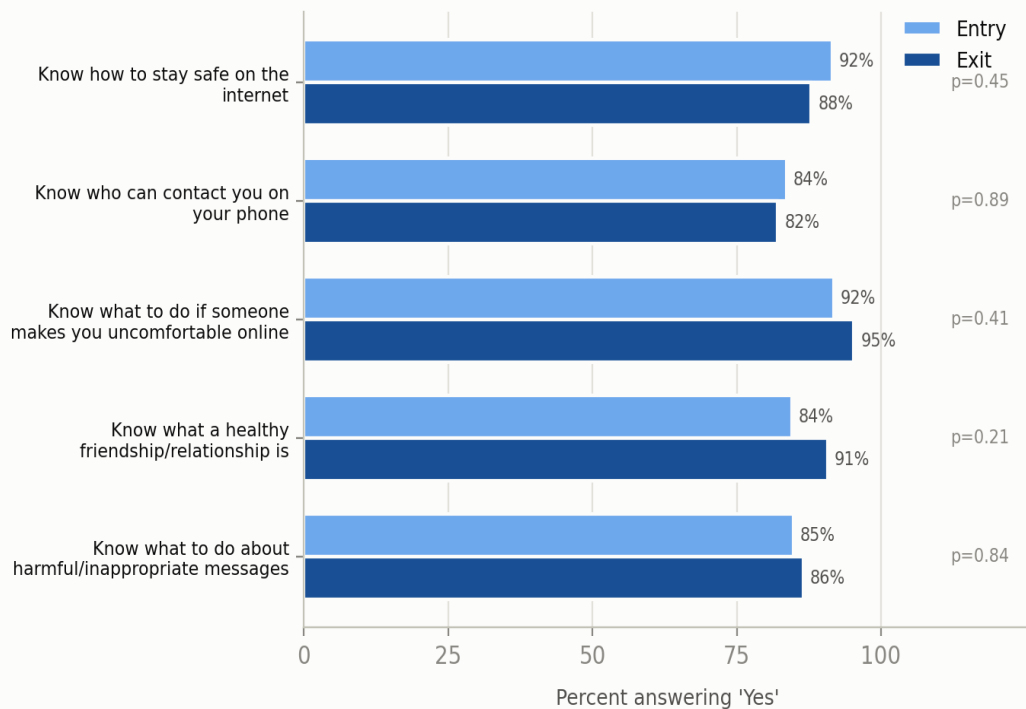


Composite means reinforce the item-level picture: program-attributed change averaged 2.01 and future intentions 1.97 on the 1-to-5 scale where 1 is much more likely, both landing near the somewhat more likely anchor. Program experience averaged 1.76 on the 1-to-4 scale where 1 is all of the time, indicating that most youth felt engaged and respected most or all of the time.

3.4 Items asked at both entry and exit: the formal pre/post test

Figure 4 and Table 2 present the five life-skills knowledge items administered identically at both waves. Entry knowledge was already high, between 84 and 92 percent Yes, and exit percentages were statistically indistinguishable from entry on every item; after false discovery rate adjustment, all p values exceed 0.75. The knowledge composite tells the same story: 0.91 at entry and 0.92 at exit on the 0-to-1 scale (Welch $t = 0.17$, $p = 0.86$, $d = 0.02$). Two items moved in the favorable direction and two moved slightly against it, a pattern consistent with sampling noise around a ceiling.

Figure 4. Life-skills knowledge asked at BOTH entry and exit



Entry n = 146-153; exit n = 106-107. No comparison significant (all FDR-adjusted $p \geq .75$); entry scores were already near ceiling.

Table 2. Pre/post tests on shared knowledge items (independent samples)

Item	Entry % Yes (n)	Exit % Yes (n)	p	FDR p	Cohen's d
Know how to stay safe on the internet	91.5% (153)	87.9% (107)	0.45	0.75	-0.17
Know who can contact you on your phone	83.6% (152)	82.1% (106)	0.89	0.89	-0.08
Know what to do if someone makes you uncomfortable online	91.8% (147)	95.3% (106)	0.41	0.75	0.13
Know what a healthy friendship/relationship is	84.5% (148)	90.7% (107)	0.21	0.75	0.13
Know what to do about harmful/inappropriate messages	84.8% (138)	86.5% (104)	0.84	0.89	0.08

Note. FDR p = the p value adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate procedure; because five knowledge items are tested at once, the chance of at least one falsely significant result rises, and the adjustment corrects each p value so that the expected proportion of false positives across the family of tests stays at or below 5 percent. Knowledge composite: entry mean 0.91 ($n = 152$), exit mean 0.92 ($n = 107$), Welch $t = 0.17$, $p = 0.8642$, $d = 0.02$. Scores: Yes = 1, Unsure = 0.5, No = 0.

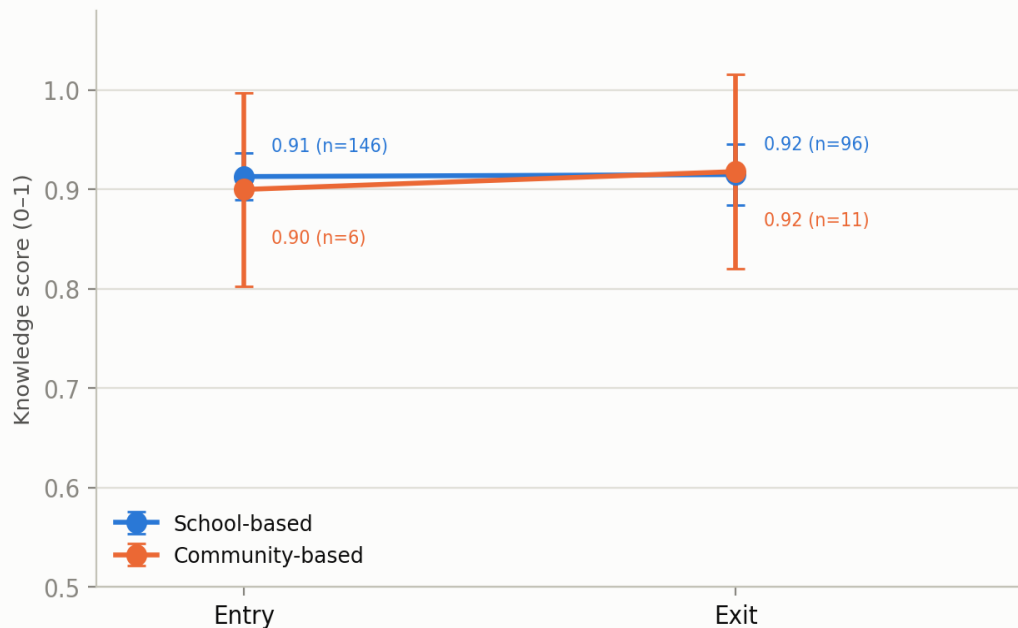
3.5 Community-based cohorts versus school-based cohorts

The community cohorts served slightly older youth (mean age 13.33 at entry versus 11.69 in school cohorts) in camp and community settings. Table 3 shows descriptives for the three item families by group, and Figure 5 plots the knowledge composite by group and time.

Table 3. Descriptives by setting

Measure	Community-based	School-based
Entry / exit surveys, n	6 / 18	190 / 153
Age at entry, mean (SD)	13.33 (0.52), n=6	11.69 (0.84), n=189
Entry: any substance use, %	33.3% (n=6)	13.4% (n=187)
Entry: skill frequency (1-4, lower better)	3.4 (0.42), n=5	2.73 (0.69), n=183
Entry: goal self-beliefs (1-3, higher better)	2.27 (0.45), n=6	2.41 (0.44), n=187
Entry: relationship confidence (1-3, higher better)	2.28 (0.65), n=6	2.47 (0.5), n=180
Knowledge composite at entry (0-1)	0.9 (0.12), n=6	0.91 (0.14), n=146
Knowledge composite at exit (0-1)	0.92 (0.17), n=11	0.92 (0.15), n=96
Exit: program-attributed change (1-5, lower better)	2.24 (1.21), n=17	1.98 (1), n=146
Exit: future intentions (1-5, lower better)	2.39 (1.17), n=14	1.92 (1.04), n=137
Exit: program experience (1-4, lower better)	1.91 (0.65), n=15	1.74 (0.75), n=134

Figure 5. Knowledge composite by group and time

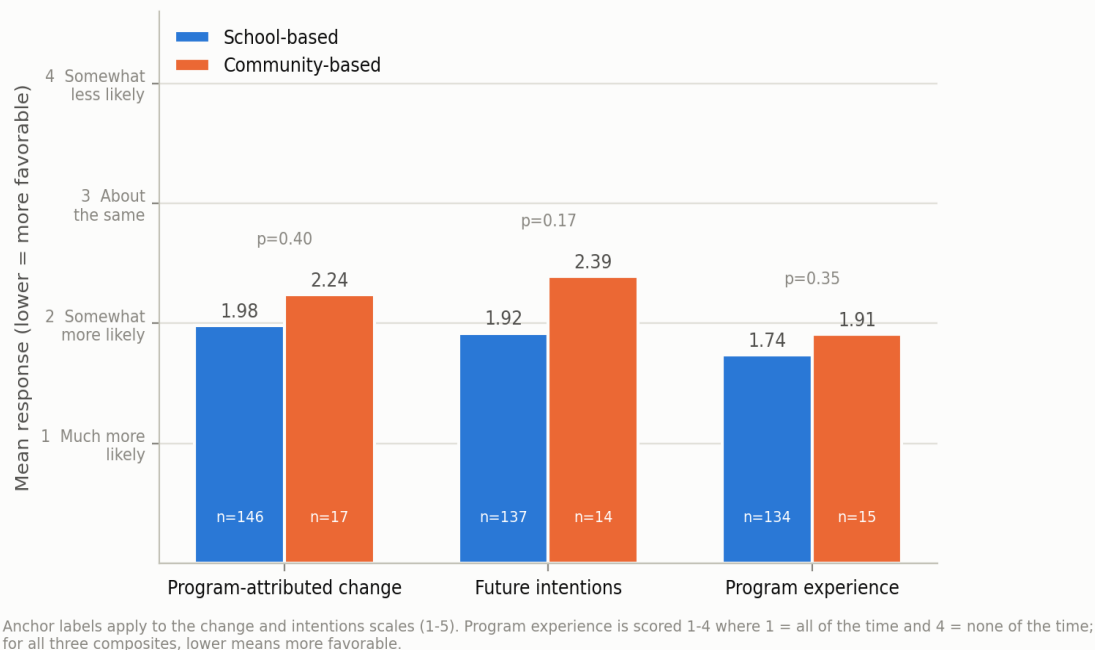


Independent samples at each time point; bars are 95% CIs. Group $\times$ time interaction: $b = +0.016$, $p = 0.84$ (not significant).

Differential change. The time by group interaction on the knowledge composite was near zero and not significant ($b = 0.0157$, $p = 0.8403$), meaning the two settings changed at statistically indistinguishable rates. Within each group the entry-to-exit difference was likewise null (community: 0.9 to 0.92, $p = 0.8012$; school-based: 0.91 to 0.92, $p = 0.9024$).

Between-group differences at exit. On every exit composite the community cohorts trended less favorable than the school cohorts (program-attributed change 2.24 versus 1.98; future intentions 2.39 versus 1.92; program experience 1.91 versus 1.74), but no difference reached significance (all p at or above 0.17), and the community samples are small (Figure 6).

Figure 6. Exit composites: community vs school-based cohorts



4. Discussion

4.1 What the data show

Three storylines emerge. The first concerns what participants brought with them: these middle schoolers arrived with substantial existing knowledge of online safety and healthy relationships, low levels of substance use, and moderate self-beliefs, but with two visible gaps, infrequent resistance to peer pressure and very little communication with parents or caregivers about important decisions. A program hoping to demonstrate knowledge gains on the five life-skills questions had little headroom to do so; when nine in ten students answer correctly at entry, even a perfectly effective curriculum cannot move the number far, and measured change will hover near zero exactly as observed. This is a measurement ceiling, a well-documented threat to sensitivity in pre/post designs (Wang et al., 2008), not evidence that the program taught nothing.

The second storyline is the participants' own account of change. On exit, roughly seven in ten youth said the program made them more likely to plan for goals, care about school, manage emotions, and understand healthy relationships, and about two in three reported strengthened intentions to delay sexual activity through high school and to sequence employment and marriage before children, which are the core SRAE outcomes. Self-attributed change is a weaker form of evidence than a measured pre/post gain because it invites social desirability and demand effects (Brenner et al., 2003; Howard, 1980), yet it is the only instrument this survey design points at the program's central constructs, and its internal pattern is coherent: the item ranked lowest by youth (talking with parents, 40 percent) matches the weakest baseline behavior (15 percent talking with parents often), which suggests respondents were discriminating

rather than uniformly agreeable. The parent-communication gap deserves programmatic attention in its own right, since parent-adolescent communication about sex is a meta-analytically supported protective factor for safer adolescent behavior (Widman et al., 2016).

The third storyline concerns setting. The three community cohorts showed no statistically distinguishable differences from the school cohorts on any tested measure. Descriptively, however, the community cohorts sat on the less favorable side of every exit composite and reported roughly double the baseline substance use (33 percent versus 13 percent, on very few respondents). With 6 entry and 18 exit surveys in the community group, the study is powered only to detect very large differences, so the correct reading is not that settings are equivalent but that nothing alarming is visible at this sample size.

4.2 Limitations

Five limitations shape what can be concluded. First, anonymity prevents within-person matching, so all pre/post tests compare overlapping but unmatched samples. Second, the entry and exit instruments measure different constructs under shared question numbers; only the five knowledge items are truly repeated. Third, the knowledge items exhibit a strong ceiling. Fourth, attrition is visible (196 entry versus 171 exit surveys) and two cohorts lack a wave entirely. Fifth, all outcome data are self-reported by middle school students in supervised settings, where socially desirable responding is plausible, particularly for the retrospective program-attribution items (Brenner et al., 2003).

A data-stewardship note belongs alongside these design limits. The cohort workbooks arrived with two different header layouts, and one cohort's internal label was inconsistent with its file name; both issues were detected and corrected during data preparation, but they signal that the pipeline from classroom to spreadsheet has room for tightening.

4.3 Recommendations

For the next evaluation cycle, five changes would materially strengthen the evidence. First, add an anonymous self-generated linking code so entry and exit surveys can be matched within person while preserving anonymity, a method with an established track record in school-based survey research (Kearney et al., 1984); this single change would convert the design from independent cross-sections to true paired pre/post measurement. Second, repeat identical outcome items at both waves, especially the skill frequency and self-belief batteries, so that change on the program's core constructs can be measured directly rather than inferred from retrospective attribution. Third, replace or augment the near-ceiling knowledge items with harder, scenario-based questions that leave room for growth. Fourth, standardize the workbook template across cohorts and validate each file at collection. Fifth, given the less favorable trend and higher baseline risk in community settings, collect larger samples there and consider brief fidelity logs to test whether implementation differences explain the setting gap.

4.4 Conclusion

Across 16 cohorts, this SRAE program reached the intended middle school population at a developmentally strategic age, achieved high engagement across schools, after-school programs, and camps, and was credited by two thirds or more of exiting youth with strengthening the exact skills and intentions the Title V SRAE program was created to build. The evaluation design, rather than the program, is what limits stronger claims: knowledge items were too easy for the population, and the instruments measured different things at entry and exit. The recommended design changes, particularly anonymous linking codes and repeated outcome batteries, would allow the next report to state findings in terms of measured change rather than participant attribution.

References

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300.
- Brener, N. D., Billy, J. O. G., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health*, 33(6), 436–457.
- Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *The Annals of the American Academy of Political and Social Science*, 591(1), 98–124.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Family and Youth Services Bureau. (2020). Title V State Sexual Risk Avoidance Education (SRAE) program fact sheet. U.S. Department of Health and Human Services, Administration for Children and Families.
- Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, 4(1), 93–106.
- Kearney, K. A., Hopkins, R. H., Mauss, A. L., & Weisheit, R. A. (1984). Self-generated identification codes for anonymous collection of longitudinal questionnaire data. *Public Opinion Quarterly*, 48(1B), 370–378.
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1), 50–60.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Social Security Act § 510, 42 U.S.C. § 710 (2018), as amended by the Bipartisan Budget Act of 2018, Pub. L. No. 115-123.
- Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2008). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behavioral Research*, 43(3), 476–496.

Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 34(1–2), 28–35.

Widman, L., Choukas-Bradley, S., Noar, S. M., Nesi, J., & Garrett, K. (2016). Parent–adolescent sexual communication and adolescent safer sex behavior: A meta-analysis. *JAMA Pediatrics*, 170(1), 52–61.

About this sample

This document is a de-identified version of a real deliverable: every statistic, figure, and table was produced from an actual grantee's 30 survey workbooks, and every client, site, curriculum, and location identifier has been removed or renumbered. A full engagement also includes the cleaned analytic dataset with a complete codebook and an Excel workbook containing every statistical table.

Wherever this sample says [Your Organization's Name Here], the real deliverable carries your branding and your authorship; you stay the evaluator of record on every page. Intellectus 360 provides the analysis, findings, conclusions, and recommendations behind it, for SRAE and PREP grantees, quantitative or qualitative, delivered within five business days of complete data.

Info@Intellectus360.com | (888) 383-6639 | Intellectus360.com