Appendix

Appendix A1 Study characteristics: Hall & Bacon, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Hall, B. W., & Bacon, T. P. (2005). Building a foundation against violence: Impact of a school-based prevention program on elementary students. *Journal of School Violence,* 4(4), 63–83.
Participants	The study included 999 third-grade students from 10 elementary schools. Of the sample, 48% were females, 20% received exceptional education services, 17% received limited English proficiency services, 44% were Caucasian, 12.5% African-American, 36% Hispanic, and 7.5% "multicultural or other race." About 54% of the students in the sample were eligible for participation in the free or reduced lunch program.
Setting	One large school district in Florida serving students from urban, suburban, and rural regions.
Intervention	The program was implemented during the first quarter of the school year. The program instructors in the intervention group delivered seven lesson units—one a week—over a seven-week period, with each lesson averaging 45 minutes in length.
Comparison	Students in the comparison group did not participate in the *Too Good for Violence* program. In addition, the comparison schools were asked to refrain from delivering any major prevention curriculum or program until the fourth quarter.
Primary outcomes and measurement	The two measures, the Student Protective Factor Survey Questionnaire and the Teacher Checklist of Student Behaviors, were administered immediately after the intervention and again 20 weeks later. (See Appendices A2.1 and A2.2.)
Teacher training	The program was delivered by program instructors (off-site educators). So no teacher training was conducted.

Appendix A2.1 Outcome measures in the behavior domain

Outcome measure	Description
Teacher Checklist of Student Behaviors (TCSB)	Teacher Checklist of Student Behaviors (TCSB). A teacher survey of 21 behavioral items rated on a five-point scale. Response options ranged from 1 (never) to 5 (almost always). The TCSB, developed by the study authors, consists of three subscales: personal and social skills; positive social behaviors; and inappropriate social behaviors (as cited in Hall & Bacon, 2005).

Appendix A2.2 Outcome measures in the knowledge, attitudes, and values domain

Outcome measure	Description
Student Protective Factors Survey Questionnaire (SPFSQ)	Student Protective Factors Survey Questionnaire (SPFSQ). A student survey of 32 items rated on a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The SPFSQ, developed by the study authors, consists of four subscales: emotional competency, social and resistance skills, communication skills, and interactions with others (as cited in Hall & Bacon, 2005).

Appendix A3.1 Summary of study findings included in the rating for the behavior domain[1]

Outcome measure	Study sample	Sample size (schools)	Author's findings from the study: Mean outcome (standard deviation[2]) — *Too Good for Violence* group (column 1)	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group (column 2)	WWC calculations: Mean difference[3] (column 1–column 2)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Hall & Bacon, 2005 (randomized controlled trial)								
Teacher Checklist of Student Behaviors (20-week follow-up)—total score	Grade 3	10	4.17 (0.16)	3.86 (0.82)	0.31	0.47	Statistically significant	+18
Domain average[7] for behavior						0.47	Statistically significant	+18

1. This appendix reports overall findings 20 weeks after the end of the program considered for the effectiveness rating and the improvement index. End-of-program findings from the same studies are not included in these ratings but are reported in Appendix A4.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations reported by the study authors for the school-level analysis represent variations among schools rather than students. The standard deviations shown in this table were estimated by the WWC to reflect variations among students.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups. The level of statistical significance was calculated by the WWC and where necessary, corrects for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of the *Too Good for Violence* report, no corrections for clustering and multiple comparisons were needed. The study also reported findings using the classroom and the student levels as the units of analysis. The classroom-level findings were statistically significant and favored the intervention group on the teacher checklist of student behavior. The student-level analysis reported statistically significant effects favoring the intervention group on two out of three subscales of the teacher checklist of student behavior. The classroom-level and student-level analyses were not reviewed because they did not account for clustering within classrooms or schools and multiple comparisons.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
7. This row provides the study average, which is also the domain average in this case. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the knowledge, attitudes, and values domain[1]

Outcome measure	Study sample	Sample size (students/ schools)	Author's findings from the study: Mean outcome (standard deviation[2]) — *Too Good for Violence* group (column 1)	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group (column 2)	WWC calculations: Mean difference[3] (column 1– column 2)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Hall & Bacon, 2005 (randomized controlled trial)								
Student Protective Factor Survey (20-week follow-up)—total score	Grade 3	10	3.89 (0.44)	3.7 (0.41)	0.19	0.40	ns	+16
Domain average[7] for knowledge, attitudes, and values						0.40	ns	+16

ns = not statistically significant

1. This appendix reports overall findings considered for the effectiveness rating and the improvement indexSubtest and subgroup findings from the same studies are not included in these ratings but are reported in Appendix A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations reported by the study authors for the school-level analysis represent variations among schools rather than students. The standard deviations shown in this table were estimated by the WWC to reflect variations among students.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups. The level of statistical significance was calculated by the WWC and where necessary, corrects for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of the *Too Good for Violence* report, no corrections for clustering and multiple comparisons were needed. The study also reported findings using the classroom and the student levels as the units of analysis. The classroom-level findings were statistically significant and favored the intervention group on the student protective factors survey. The student-level analysis reported statistically significant effects favoring the intervention group on three out of four subscales of the student protective factors survey. The classroom-level and student-level analyses were not reviewed because they did not account for clustering within classrooms or schools and multiple comparisons.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
7. This row provides the study average, which is also the domain average in this case. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Summary of end-of-program study findings for the behavior domain[1]

Outcome measure	Study sample	Sample size (students/ schools)	Author's findings from the study: Mean outcome (standard deviation[2]) — *Too Good for Violence* group (column 1)	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group (column 2)	WWC calculations: Mean difference[3] (column 1– column 2)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Hall & Bacon, 2005 (randomized controlled trial)								
Teacher Checklist of Student Behaviors (end of program)—total score	Grade 3	10	4.18 (0.47)	3.87 (0.66)	0.31	0.49	Statistically significant	+19

1. This appendix presents end-of-program findings for the behavior domain. Follow-ups 20 weeks after the end of the program were used for rating purposes and are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations reported by the study authors for the school-level analysis represent variations among schools rather than students. The standard deviations shown in this table were estimated by the WWC to reflect variations among students.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups. The level of statistical significance was calculated by the WWC and where necessary, corrects for clustering within classrooms or schools. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of the *Too Good for Violence* report, no corrections for clustering were needed. The study also reported findings using the classroom and the student levels as the units of analysis. The classroom-level findings were statistically significant and favored the intervention group on the teacher checklist of student behavior. The student-level analysis reported statistically significant effects favoring the intervention group on two out of three subscales of the teacher checklist of student behavior. The classroom-level and student-level analyses were not reviewed because they did not account for clustering within classrooms or schools and multiple comparisons.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.

Appendix A4.2 Summary of end-of-program study findings for the knowledge, attitudes, and values domain[1]

Outcome measure	Study sample	Sample size (schools)	Author's findings from the study: Mean outcome (standard deviation[2]) — *Too Good for Violence* group (column 1)	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group (column 2)	WWC calculations: Mean difference[3] (column 1–column 2)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Hall & Bacon, 2005 (randomized controlled trial)								
Student Protective Factor Survey (end of program)—total score	Grade 3	10	4.04 (0.60)	3.82 (0.60)	0.22	0.33	ns	+13

ns = not statistically significant

1. This appendix presents end-of-program findings for measures in the knowledge, attitudes, and values domain. Follow-ups 20 weeks after the end of the program were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations reported by the study authors for the school-level analysis represent variations among schools rather than students. The standard deviations shown in this table were estimated by the WWC to reflect variations among students.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups. The level of statistical significance was calculated by the WWC and where necessary, corrects for clustering within classrooms or schools. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of the *Too Good for Violence* report, no corrections for clustering were needed. The study also reported findings using the classroom and the student levels as the units of analysis. The classroom-level findings were statistically significant and favored the intervention group on the student protective factors survey. The student-level analysis reported statistically significant effects favoring the intervention group on three out of four subscales of the student protective factors survey. The classroom-level and student-level analyses were not reviewed because they did not account for clustering within classrooms or schools and multiple comparisons.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.

Appendix A5.1 Rating for the behavior domain

The WWC rates interventions as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of behavior, the WWC rated *Too Good for Violence* as having potentially positive effects. It did not meet the criteria for positive effects, because it only had one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Too Good for Violence* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, thus qualifying as a *positive* effect.

  Met. The one study on *Too Good for Violence* had a positive and statistically significant effect size for the single outcome in this domain.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect. Fewer or the same number of studies showing indeterminate effects than showing statistically significant or substantively important positive effects.

  Met. The WWC analysis found no indeterminate effects or statistically significant substantively important negative effects in this domain.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. *Too Good for Violence* had only one study meeting WWC evidence standards for a strong design.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The one study meeting WWC evidence standards did not show statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.2 Rating for the knowledge, attitudes, and values domain

The WWC rates interventions as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of knowledge, attitudes, and values, the WWC rated *Too Good for Violence* as having potentially positive effects. It did not meet the criteria for positive effects, because it had only one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Too Good for Violence* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, thus qualifying as a *positive* effect.

  Met. The one study on *Too Good for Violence* in this domain had a substantially important effect size (greater than 0.25), which the WWC considers a positive effect.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect. Fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. The WWC analysis found no indeterminate effects or statistically significant or substantively important negative effects in this domain.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. *Too Good for Violence* had only one study meeting WWC evidence standards for a strong design.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The one study meeting WWC evidence standards did not show statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive effects. See the WWC Intervention Rating Scheme for a complete description.