

Appendix

Appendix A1 Study Characteristics: Grossman and Sipe, 1992 (randomized controlled trial)

Characteristic	Description
Study citation	Grossman, J. B., & Sipe, C. L. (1992). Summer Training and Education Program (STEP): <i>Report on long-term impacts</i> . Philadelphia, PA: Public/Private Ventures.
Participants	<p>The study used a randomized controlled trial research design to examine the effects of <i>STEP</i> in five sites located in four states. The sample included youth who had applied for <i>STEP</i> in 1986 and 1987 in five research sites. To be eligible, youth had to be 14 or 15 years old and from low-income families and had to have tested below grade level.</p> <p>The study included three cohorts of youth. The authors analyzed each of these cohorts separately. The Cohort 1 analysis did not meet WWC evidence standards because of high attrition and a failure to establish baseline equivalence. This analysis is therefore not included in this report. The Cohort 2 and Cohort 3 analyses meet evidence standards. Since these analyses were conducted in the same sites, the WWC treated these analyses as one study and combined the results when rating the effectiveness of <i>STEP</i>.</p> <p>The Cohort 2 analysis included 1,635 eligible youth who were randomly assigned to either a treatment group, which was offered the opportunity of participating in the <i>STEP</i> program during the summers of 1986 and 1987, or to a control group, which was offered a summer job in the federally funded Summer Youth Employment and Training Program (SYETP). The study lost 382 students to attrition or survey nonresponse, leaving an analysis sample of 1,253 youth.</p> <p>The Cohort 3 analysis included 1,591 eligible youth who were randomly assigned to either a treatment group, which was offered the opportunity of participating in the <i>STEP</i> program during the summers of 1987 and 1988, or to a control group, which was offered a summer job in the federally funded SYETP. The study lost 256 students to attrition or survey nonresponse, leaving an analysis sample of 1,335 youth.</p> <p>Across these two cohorts, slightly more than half of the study sample was female. About half of the sample was African-American, one-fifth was Asian, and one-fifth was Hispanic. All students came from low-income households, and about half of the sample came from female-headed households. About one-third of students reported having repeated a grade, and average test scores indicated that sample students were performing substantially below their grade level in math.</p>
Setting	The five study sites were Job Training and Partnership Act local employment and training agencies that operated both <i>STEP</i> and SYETP. These sites were located in Boston, MA; Fresno, CA; Portland, OR; San Diego, CA; and Seattle, WA. Remedial education and life skills learning classes were typically held at a community school or college. Work experience was conducted in the community near where the classes were held.
Intervention	The <i>STEP</i> program focused on four areas: remediation, life skills and opportunities (LSO), work experience, and school-year support. During each of two summers, participants were offered approximately 90 hours of remedial instruction in basic reading and math skills, 18 hours of LSO instruction, and at least 80 hours of work experience. The remediation component provided a minimum of 90 hours of skill-based group and individually paced instruction. The LSO component stressed responsible social and sexual attitudes and behaviors. In the work experience component, participants were usually assigned to jobs near their remediation site. These jobs were typically in maintenance, recreation, clerical, and child care aide positions. During the intervening school year, program youth interacted with a designated counselor/advocate. These counselors helped monitor school attendance and performance of students and referred them to social services as needed.
Comparison	The control group members were offered either a one-summer or two-summer job in the federally funded SYETP. The counterfactual condition varied across sites. In general, members of the control group were provided a part-time job during the first summer for which they were eligible for the program. In two sites, San Diego and Seattle, control group youth also were provided a job in the second summer. In general, control group youth spent more time working than did the treatment group youth, since treatment group youth were receiving remedial education and life skills training in addition to work. Overall numbers of hours engaged in study-related activity appear to be roughly the same for treatment and control group youth, with control group youth participating in employment for an average of 190 hours.

(continued)

Appendix A1 Study Characteristics: Grossman and Sipe, 1992 (randomized controlled trial) (continued)

Characteristic	Description
Primary outcomes and measurement	The relevant study outcomes included in this review are whether students dropped out of school and their highest grade completed, based on student follow-up interviews. For a more detailed description of these outcome measures, see Appendices A2.1 and A2.2.
Staff/teacher training	<p>Each site hired a lead teacher who was primarily responsible for overseeing the remediation component, hiring teachers, planning and providing teacher training, and assembling a curriculum development team of reading and math teachers to develop curriculum modules for site use. The lead teacher also assumed responsibility for seeing that life skills staff were supported and successfully integrated into program operations. Remediation teachers also received in-service training over the course of the summer program, an average of 4 to 6 hours total. In addition, before this cohort's second summer of participation, the evaluator's consultants and staff delivered preservice training to all remediation teachers at participating sites. <i>STEP</i> teachers received an average of between 15 and 20 hours of preservice training (the hours varied by site). In addition to specific topics outlined in the training guides, teacher training included an orientation to the program, discussions on the integration of the life skills and work experience components, and a review of modules and program logistics.</p> <p>For the life skills component, all life skills instructors attended a two-day comprehensive training workshop. This training included a review of curriculum content, an analysis of instructional techniques, a discussion of classroom management issues, and a series of role-plays and other sensitizing activities. They received minimal in-service training.</p>

Appendix A2.1 Outcome measures for the staying in school domain

Outcome measure	Description
Dropped out	Whether the student reported having dropped out of school. These self-reported data were collected from follow-up surveys. For Cohort 2, the follow-up survey was conducted four-and-one-half years after random assignment. For Cohort 3, the follow-up survey was conducted three-and-one-half years after random assignment.

Appendix A2.2 Outcome measures for the progressing in school domain

Outcome measure	Description
Highest grade completed	Number of years of school completed. These self-reported data were collected from follow-up surveys. For Cohort 2, the follow-up survey was conducted four-and-one-half years after random assignment. For Cohort 3, the follow-up survey was conducted three-and-one-half years after random assignment.

Appendix A3.1 Summary of study findings included in the rating for the staying in school domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome		Mean difference ² (STEP—comparison)	Effect size ³	Statistical significance ⁴ (at $\alpha = 0.05$)	Improvement index ⁵
			STEP group	Comparison group				
Grossman and Sipe, 1992 (randomized controlled trial)⁶								
Dropped out	Cohort 2	1,253	29.2	25.1	−4.1	−0.13	ns	−7
Dropped out	Cohort 3	1,335	21.6	22.6	1.0	0.04	ns	2
Domain average for staying in school across study findings⁷						−0.06	ns	−2

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the staying in school domain. The rating for the staying in school domain is based only on findings for dropout measures from student follow-up surveys. Findings for dropout measures from school reports are not included because the study authors argue that these measures are inaccurate.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For the dropout outcome, signs were reversed on the mean difference, effect size, and improvement index to demonstrate that the treatment group was favored when negative differences were reported. Mean differences reflect regression-adjusted treatment impacts estimated by the authors. Mean outcomes for the STEP group were calculated by adding this impact estimate to the comparison group mean.
3. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
5. 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.
6. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. For the STEP study summarized here, no corrections for clustering or multiple comparisons were needed.
7. The study also included a Cohort 1 analysis that did not meet WWC evidence standards. These results were therefore not used to rate the effectiveness of STEP.

Appendix A3.2 Summary of study findings included in the rating for the progressing in school domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome		Mean difference ² (STEP—comparison)	Effect size ³	Statistical significance ⁴ (at $\alpha = 0.05$)	Improvement index ⁵
			STEP group	Comparison group				
Grossman and Sipe, 1992 (randomized controlled trial)⁶								
Highest grade completed	Cohort 2	1,231	11.2	11.2	0.0	n/a	ns	n/a
Highest grade completed	Cohort 3	1,327	10.8	10.9	0.0	n/a	ns	n/a
Domain average for progressing in school across all studies⁷						n/a	ns	n/a

ns = not statistically significant

n/a = not available⁸

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the progressing in school domain. The rating for the progressing in school domain is based only on findings for highest grade completed measures from student follow-up surveys. Findings for dropout measures from school reports are not included because the study authors argue that these measures are inaccurate.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. Mean differences reflect regression-adjusted treatment impacts estimated by the authors. Mean outcomes for the *STEP* group were calculated by adding this impact estimate to the comparison group mean.
3. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
5. 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.
6. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. For the *STEP* study summarized here, no corrections for clustering or multiple comparisons were needed.
7. The study also included a Cohort 1 analysis that did not meet WWC evidence standards. These results were therefore not used to rate the effectiveness of *STEP*.
8. The WWC was unable to calculate an effect size or an improvement index in the progressing in school domain because the study authors did not report a standard deviation for the outcome.

Appendix A4 Summary of graduation rate findings for the completing school domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study					
			Mean outcome		WWC calculations			
			STEP group	Comparison group	Mean difference ² (STEP—comparison)	Effect size ³	Statistical significance ⁴ (at $\alpha = 0.05$)	Improvement index ⁵
Grossman and Sipe, 1992 (randomized controlled trial)⁶								
Earned a high school diploma or GED	Students old enough to graduate in Cohort 2	1,119	58.1	63.9	–5.8	–0.20	nr	–8
Earned a high school diploma or GED	Students old enough to graduate in Cohort 3	539	64.8	62.4	2.4	0.08	ns	3

ns = not statistically significant

nr = not reported

1. This appendix presents findings for measures that fall in the completing school domain. The outcome was collected from student follow-up surveys and represents whether the student had received a high school diploma or GED. For Cohort 2, the follow-up survey was conducted four-and-one-half years after random assignment. For Cohort 3, the follow-up survey was conducted three-and-one-half years after random assignment. The Cohort 2 findings were not used for rating purposes because the analysis sample exhibited high attrition and the study did not present evidence that the research groups from the analysis sample were equivalent at baseline. The Cohort 3 findings were not used for rating purposes because the study did not provide sufficient information to calculate the rate of attrition for the analysis sample, nor did it present evidence that the research groups from the analysis sample were equivalent at baseline.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
3. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. Authors reported statistical significance for graduation rates at the $\alpha = 0.10$ level, so it was not possible to determine whether findings were significant at the $\alpha = 0.05$ level.
5. 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.
6. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. For the STEP study summarized here, no corrections for clustering or multiple comparisons were needed.

Appendix A5.1 STEP rating for the staying in school domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹ For the outcome domain of staying in school, the WWC rated *STEP* as having no discernible effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

Met. No studies of *STEP* showed a statistically significant or substantively important effect.

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. No studies of *STEP* showed a statistically significant or substantively important effect.

AND

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

Met. No studies of *STEP* showed a statistically significant or substantively important negative effect.

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.

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

AND

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

Met. No studies of *STEP* showed a statistically significant or substantively important effect.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

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 or potentially negative effects. For a complete description, see the WWC Intervention Rating Scheme.

(continued)

Appendix A5.1 *STEP* rating for the staying in school domain *(continued)*

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

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies of *STEP* showed a statistically significant or substantively important effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

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

Met. No studies of *STEP* showed a statistically significant or substantively important effect.

Appendix A5.2 STEP rating for the progressing in school domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹ For the outcome domain of progressing in school, the WWC rated *STEP* as having no discernible effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

Met. No studies of *STEP* showed a statistically significant or substantively important effect.

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. No studies of *STEP* showed a statistically significant or substantively important effect.

AND

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

Met. No studies of *STEP* showed a statistically significant or substantively important effect.

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.

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

AND

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

Met. No studies of *STEP* showed a statistically significant or substantively important effect.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

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 or potentially negative effects. For a complete description, see the WWC Intervention Rating Scheme.

(continued)

Appendix A5.2 STEP rating for the progressing in school domain *(continued)*

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

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies of *STEP* showed a statistically significant or substantively important effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence..

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

Not met. No studies of *STEP* showed a statistically significant or substantively important effect.

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

Met. No studies of *STEP* showed a statistically significant or substantively important effect.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Staying in school	1	>2	2,588	Small
Progressing in school	1	>2	2,558	Small

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.”