Appendix

Appendix A1 Study characteristics: Leming, 2001 (quasi-experimental design)

Characteristic	Description
Study citation	Leming, J. S. (2001). Integrating a structured ethical reflection curriculum into high school community service experiences: Impact on students' sociomoral development. *Adolescence,* 36, 33–45.
Participants	The study involved 283 twelfth-grade students in one high school. The school district was described as serving middle-class communities with a mainly white population. Students were described as college preparatory, and the sample was 75% female and 25% male.
Setting	A suburban high school in St. Louis, Missouri.
Intervention	The intervention group participated in a community service elective course that implemented the *Building Decision Skills* curriculum. Class instruction had the students working in large and small groups with homework assigned after each of 10 lessons. The lessons were taught during the first two weeks of the semester. The community service component, integrated with the *Building Decision Skills* curriculum, involved two days a week off-campus and two days on-campus participating in school and community service activities, such as providing companionship to residents of a retirement home or running a recycling center on campus.
Comparison	The comparison group was drawn from senior English literature classes at the same high school as the intervention group. Comparison group students did not participate in *Building Decision Skills* or the community service course.
Primary outcomes and measurement	Students responded to a study-specific questionnaire that included three ethical dilemmas for which their responses were scored on ethical awareness, ethical responsibility, and ethical perspective. Additional measures were used to assess students' self-esteem in social settings (Janis-Field Feelings of Inadequacy Scale), their sense of social responsibility in school and in society, and their anticipation of future community participation (Newmann and Rutter's Moral-Political Awareness Scale). (See Appendix A2.)
Teacher training	No information on teacher training was provided.

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

Outcome measure	Description
Ethical awareness	One of three measures developed by the researcher based on student responses to three scenarios presenting ethical dilemmas. This measure assesses the respondent's recognition of an ethical dimension in the scenarios. For this measure the WWC averaged scores over the three scenarios used by the researcher (as cited in Leming, 2001).
Ethical responsibility	One of three measures developed by the researcher based on student responses to three scenarios presenting ethical dilemmas. This measure assesses the respondent's views on who has responsibility for different decisions (self or others). For this measure the WWC averaged scores over the three scenarios used by the researcher (as cited in Leming, 2001).
Ethical perspective	One of three measures developed by the researcher based on student responses to three scenarios presenting ethical dilemmas. This measure assesses the extent to which the respondent frames issues from perspectives consistent with the *Building Decision Skills* curriculum (truth versus loyalty). For this measure the WWC averaged scores over the three scenarios used by the researcher (as cited in Leming, 2001).
Janis-Field Feelings of Inadequacy scale	The Janis-Field Feelings of Inadequacy scale assessed students' sense of self-esteem in social settings after doing community service (as cited in Leming, 2001).
Newmann and Rutter's Moral-Political Awareness scale, Social responsibility (general) subscale	A seven-item subscale ranging from 7 to 35, originally developed by Newmann and Rutter, measuring sense of responsibility and concern for the welfare of others regarding society at large (as cited in Leming, 2001).
Newmann and Rutter's Moral-Political Awareness scale, Social responsibility (school) subscale	A three-item subscale ranging from 3 to 15, originally developed by Newmann and Rutter, measuring sense of responsibility and concern for the welfare of others in school (as cited in Leming, 2001).
Newmann and Rutter's Moral-Political Awareness scale, Anticipated community participation subscale	An adapted version of a subscale developed by Newmann and Rutter to measure the disposition of students toward greater community involvement (as cited in Leming, 2001).

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

Outcome measure	Study sample	Sample size[e] (classrooms/ students)	Author's findings from the study: Mean outcome[b] (standard deviation[c]) *Building Decision Skills* group (column 1)	Author's findings from the study: Mean outcome[b] (standard deviation[c]) Comparison group (column 2)	WWC calculations: Mean difference[f] (column 1– column 2)	WWC calculations: Effect size[g]	WWC calculations: Statistical significance[h] (at α = 0.05)	WWC calculations: Improvement index[i]
Leming, 2001 (quasi-experimental design)								
Ethical awareness[j]	Grade 12	6/283	2.25 (na)	1.85 (na)	0.40	0.56	ns	+21
Ethical responsibility[j]	Grade 12	6/283	2.22 (na)	1.97 (na)	0.25	0.33	ns	+13
Ethical perspective[j]	Grade 12	6/283	1.84 (na)	1.31 (na)	0.53	0.84	Statistically significant	+30
Self-esteem	Grade 12	6/279	35.03 (5.76)	34.57 (6.44)	0.46	0.07	ns	+3
Social responsibility (general)	Grade 12	6/282	26.52 (3.18)	26.16 (4.13)	0.36	0.09	ns	+4
Social responsibility (school)	Grade 12	6/282	10.89 (2.13)	9.70 (3.06)	1.19	0.42	ns	+17
Anticipated community participation	Grade 12	6/278	16.19 (2.76)	15.52 (3.16)	0.67	0.22	ns	+9
Domain average[k] for behavior						0.36	ns	+14

ns = not statistically significant

a. This appendix reports summary findings considered for the effectiveness rating and the improvement index. The WWC averaged individual items to provide the scales reported here for three outcomes (ethical awareness, ethical responsibility, and ethical perspective). Findings on the individual item level are presented in Appendix A4.
b. The study author adjusted the means for pretest differences between groups.
c. 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 for all outcomes were requested by the WWC and submitted by the study author.
d. Leming (2001) also examined effects on students' rankings of social values, but this outcome was not included in the WWC review because the report did not specify which of the social values the curriculum was expected to change relative to other values. Without that information, no determination could be made about whether the results reported were favorable or unfavorable. In addition, severe attrition of respondents between time of pretest and time of posttest for this measure prevents the WWC from making valid conclusions about this outcome. The study also included a comparison between the *Building Decision Skills* curriculum combined with service-learning and another condition involving service-learning only. In that comparison, differences favoring the curriculum plus service-learning were reported for ethical awareness and ethical perspective in comparison with service-learning alone. Differences were not found for acceptance of ethical responsibility, social responsibility within the school, social responsibility in society, self-esteem, or anticipated future community participation. The WWC Character Education review looks at effects of character education programs compared to "business as usual," or no program. The analysis that compares BDS (with service-learning) to another condition (service-learning alone) does not have a "business as usual" condition and so is not within the focus of this review.
e. The six classrooms that participated in this study represent three cohorts of students (two classes per semester for three consecutive semesters).
f. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
g. For an explanation of the effect size calculation, please see the WWC Technical Working Paper on Effect Size.
h. 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, see the WWC Tutorial on Mismatch. See the WWC Intervention Rating Scheme for the formulas the WWC used to calculate the statistical significance. In the case of the *Building Decision Skills* report, corrections for clustering and multiple comparisons were needed.
i. 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.
j. This outcome measure is the average of three items reported by the study author. Means were computed as simple averages of the item means reported by the study author. Standard deviations across the three items could not be computed. Effect sizes were computed at the item level and then averaged. Table A4 presents summary characteristics on the item level.
k. 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 Summary of detailed study findings in the knowledge, attitudes, and values domain[a]

Outcome measure: Item level	Study sample	Sample size (classrooms/ students)	Author's findings from the study: Mean outcome[b] (standard deviation[c]) *Building Decision Skills* group (column 1)	Author's findings from the study: Mean outcome[b] (standard deviation[c]) Comparison group (column 2)	WWC calculations: Mean difference[d] (column 1– column 2)	WWC calculations: WWC-computed effect size[e]	WWC calculations: Statistical significance[f] (at α = 0.05)	WWC calculations: Improvement index[g]
Leming, 2001 (quasi-experimental design)								
Ethical Awareness								
Dilemma 1	Grade 12	6/283	2.52 (0.62)	2.09 (0.81)	0.43	0.58	ns	+22
Dilemma 2	Grade 12	6/283	2.13 (0.73)	1.75 (0.72)	0.38	0.52	ns	+20
Dilemma 3	Grade 12	6/279	2.10 (0.58)	1.72 (0.71)	0.39	0.58	Statistically significant	+22
Ethical Responsibility								
Dilemma 1	Grade 12	6/283	2.39 (0.63)	2.15 (0.82)	0.24	0.32	ns	+12
Dilemma 2	Grade 12	6/283	1.89 (0.79)	1.65 (1.64)	0.24	0.34	ns	+13
Dilemma 3	Grade 12	6/283	2.38 (0.62)	2.11 (0.91)	0.27	0.34	ns	+13
Ethical Perspective								
Dilemma 1	Grade 12	6/283	2.17 (0.76)	1.41 (0.57)	0.75	1.16	Statistically significant	+38
Dilemma 2	Grade 12	6/283	1.64 (0.65)	1.33 (0.51)	0.31	0.54	ns	+21
Dilemma 3	Grade 12	6/283	1.71 (0.77)	1.19 (0.54)	0.52	0.82	Statistically significant	+29

ns = not statistically significant

a. This appendix presents item-level findings for three scales in the knowledge, attitudes, and values domain: ethical awareness, ethical responsibility, and ethical perspective. Aggregated scale scores used for rating purposes are presented in Appendix A3.

b. The study author adjusted the means for pretest differences between groups.

c. 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.

d. Positive differences and effect sizes favor the intervention group.

e. For an explanation of the effect size calculation, please see the WWC Technical Working Paper on Effect Size.

f. 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, see the WWC Tutorial on Mismatch. See the WWC Intervention Rating Scheme for the formulas the WWC used to calculate the statistical significance. In the case of the *Building Decision Skills* report, a correction for clustering was needed.

g. 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 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 *Building Decision Skills* combined with service learning 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, negative effects) were not considered because *Building Decision Skills* 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. In the one study on *Building Decision Skills* that examined student knowledge, attitudes, and values outcomes, the average effect size was positive and substantively important. Further, the effect on one student outcome was positive and statistically significant.

- 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. No studies showed a statistically significant or substantively important negative effect. Because one study showed statistically significant positive effects and no studies showed indeterminate effects, *Building Decision Skills* met this criterion.

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. *Building Decision Skills* had only one evaluation study meeting WWC evidence standards that reported findings on knowledge, attitudes, and values, and so did not meet this criterion. Further, that study did not meet WWC evidence standards for a strong design, because it used a QED rather than an RCT design.

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

  Met. The WWC analysis found no 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.