Appendix

Appendix A1 Study characteristics: Leming, 2000 (quasi-experimental design)

Characteristic	Description
Study citation	Leming, J. S. (2000). Tell me a story: An evaluation of a literature-based character education programme. *Journal of Moral Education,* 29, 413–427.
Participants	The study included 965 students in grades 1–6 from four schools in two school districts. The student population in one school district was 95.6% Caucasian, with 25% receiving free or reduced-price lunch. The student population in the two participating schools in the other school district was 85.6% Caucasian, with 50% receiving free or reduced-price lunch.
Setting	Two participating schools (one intervention, one comparison) were in a mid-sized, semirural school district in western Pennsylvania; the other two participating schools (one intervention, one comparison) were in a small, semirural school district in southern Illinois. The comparison students were drawn from matched schools in the same school districts as the intervention students.
Intervention	The intervention group participated in *An Ethics Curriculum for Children* for the course of an academic school year. The literature-based program used folktales, folklore, and fairy tales to teach seven core values. Lessons included an opening discussion, story, discussion questions, reinforcement activity, and writing about the concept.
Comparison	No information was provided on comparison students other than they did not receive *Heartwood Ethics Curriculum for Children*.
Primary outcomes and measurement	The study-specific measures addressed understanding of the curriculum's seven core values (ethical understanding), preferences for exemplifying core values (ethical sensibility), behavior related to core values (ethical conduct), and affinity for ethnically diverse children (ethnocentrism). The measures were developed for this study to measure specific outcomes. (See Appendices A2.1 and A2.2.)
Teacher training	Teachers implementing *An Ethics Curriculum for Children* attended a half-day training course before the start of the school year.

Appendix A2.1 Outcome measures in the behavior domain

Outcome measure	Description
Ethical conduct	Ratings by teachers of character-related behavior for individual students. This measure uses a five-point Likert-type scale developed by the author (as cited in Leming, 2000).

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

Outcome measure	Description
Ethical understanding	Responses to questionnaire items developed by the author about the meaning of the curriculum's seven core attributes: courage, honesty, hope, justice, love, loyalty, and respect. Grades 1–3 used 14 items in multiple choice format (two possible answers); grades 4–6 used a sentence completion format for testing understanding of the seven attributes (as cited in Leming, 2000).
Ethical sensibility	Responses to questionnaire items developed by the author that asked about agreement with or valuing of the curriculum's seven core attributes: courage, honesty, hope, justice, love, loyalty, and respect. Grades 1–3 used a test of the student's agreement with 20 "I would" statements (e.g., "I would tell them to stop") with three levels of response (yes, no, not sure). Grades 4–6 used the same 20 items with a five-level Likert-type scale (as cited in Leming, 2000).
Ethnocentrism	Ratings of social distance, measured by photographs of ethnically diverse children for whether the student would like to have them as a friend; used with white students only. This measure was adapted from the Social Distance Measure (as cited in Leming, 2000).

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

Outcome measure[4]	Study sample	Sample size (classrooms/ students)	Author's findings from the study: Mean outcome[2] (standard deviation[3]): *Ethics Curriculum for Children* group (column 1)	Author's findings from the study: Mean outcome[2] (standard deviation[3]): Comparison group (column 2)	WWC calculations: Mean difference[5] (column 1– column 2)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at α = 0.05)	WWC calculations: Improvement index[8]
Leming, 2000 (quasi-experimental design)								
Ethical conduct	Grades 1–3	4/485	67.93 (11.54)	69.09 (14.10)	–1.16	–0.09	ns	–4
Average[9] for behavior (Grades 1–3)						–0.09	ns	–4
Ethical conduct	Grades 4–6	4/385	64.48 (12.79)	63.20 (12.54)	+1.28	0.10	ns	+4
Average[9] for behavior (Grades 4–6)						0.10	ns	+4
Domain average[9] for behavior across all studies						0.01	na	0

ns=not statistically significant
na=not applicable

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
2. The study author adjusted the means for pretest differences between groups.
3. 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.
4. The study author analyzed two additional outcomes in this domain (cheating on drawing task grades 1–3, cheating on drawing task grades 4–6) but did not provide means or standard deviations, so they are not listed in this table. None of the three outcomes was statistically significant.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
7. 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 Technical Details of WWC-Conducted Coomputations for the formulas the WWC used to calculate the statistical significance. In the case of the *Ethics Curriculum* report, a correction for clustering and multiple comparisons was needed.
8. 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.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

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

Outcome measure[4]	Study sample	Sample size (classrooms/ students)	Author's findings from the study: Mean outcome[2] (standard deviation[3]): *Ethics Curriculum for Children* group (column 1)	Author's findings from the study: Mean outcome[2] (standard deviation[3]): Comparison group (column 2)	WWC calculations: Mean difference[5] (column 1–column 2)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Leming, 2000 (quasi-experimental design)								
Ethical understanding	Grades 1–3	4/482	23.35 (2.01)	22.75 (1.93)	0.60	0.31	ns	+12
Ethnocentrism	Grades 1–3	4/447	28.10 (24.58)	33.97 (25.54)	5.87	0.23	ns	+9
Ethical sensibility	Grades 1–3	4/479	51.84 (6.87)	52.18 (6.34)	–0.34	–0.05	ns	–2
Average[9] for knowledge, attitudes, values (Grades 1–3)						0.16	ns	+6
Ethical sensibility	Grades 4–6	4/369	75.09 (11.57)	77.30 (10.30)	2.21	0.20	ns	+14
Ethical understanding	Grades 4–6	4/376	14.41 (2.44)	13.54 (2.57)	–0.87	–0.35	ns	–8
Average[9] for knowledge, attitudes, values (Grades 4–6)						0.07	ns	+3
Domain average[9] for knowledge, attitudes, values across all studies						0.12	na	+5

ns = not statistically significant
na=not applicable

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
2. The study author adjusted the means for pretest differences between groups except for one outcome (ethical understanding for grades 4–6).
3. 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.
4. The study author analyzed one additional outcome in this domain (ethnocentrism grades 4–6) but did not provide means or standard deviations, so it is not listed in this table. The outcome was not statistically significant.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The sign of the mean difference for ethnocentrism scores was reversed because greater ethnocentrism was considered a negative outcome.
6. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
7. 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 Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of the *Ethics Curriculum* report, a correction for clustering and multiple comparisons was needed.
8. 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.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A4.1 Rating for the behavior 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 behavior, the WWC rated *An Ethics Curriculum for Children* as having no discernible effects. It did not meet the criteria for positive effects because it only had one study. In addition, it did not meet the criteria for other ratings (potentially positive effects, mixed effects, potentially negative effects, and negative effects) because none of the studies showed statistically significant or substantively important effects.

Rating received

No discernible effects: No affirmative evidence of effects.

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

  Met. The WWC analysis found no statistically significant or substantively important 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. The WWC analysis found no statistically significant positive effects in this domain. *An Ethics Curriculum for Children* had only one evaluation study meeting WWC evidence standards that reported findings on behavior, and so did not meet this criterion. Further, that study did not meet WWC evidence standards for a strong 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.

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.

  Not met. The WWC analysis found no statistically significant or substantively important positive effects 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.

  Not met. No studies showed a statistically significant or substantively important negative effect. Because one study showed indeterminate effects and no studies showed statistically significant or substantively important positive effects, *An Ethics Curriculum for Children* did not meet this criterion.

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.

(continued)

Appendix A4.1 Rating for the behavior domain *(continued)*

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

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect. 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. The WWC analysis found no statistically significant or substantively important positive or negative effects in this domain.

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

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

- 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. The WWC analysis found no statistically significant or substantively important positive effects in this domain.

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 is based on a strong design.

  Not met. The WWC analysis found no statistically significant negative effects in this domain.

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

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

Appendix A4.2 Rating for the knowledge, attitudes, and values 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 knowledge, attitudes, and values, the WWC rated *An Ethics Curriculum for Children* as having no discernible effects. It did not meet the criteria for positive effects because it only had one study. In addition, it did not meet the criteria for other ratings (potentially positive effects, mixed effects, potentially negative effects, and negative effects) because none of the studies showed statistically significant or substantively important 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. The WWC analysis found no statistically significant or substantively important 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. The WWC analysis found no statistically significant positive effects in this domain. *An Ethics Curriculum for Children* 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.

- 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.

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.

  Not met. The WWC analysis found no statistically significant or substantively important positive effects 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.

  Not met. No studies showed a statistically significant or substantively important negative effect. Because one study showed indeterminate effects and no studies showed statistically significant or substantively important positive effects, *An Ethics Curriculum for Children* did not meet this criterion.

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.

(continued)

Appendix A4.2 Rating for the knowledge, attitudes, and values domain *(continued)*

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

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect. 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. The WWC analysis found no statistically significant or substantively important positive or negative effects in this domain.

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

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

- 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. The WWC analysis found no statistically significant or substantively important positive effects in this domain.

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 is based on a strong design.

  Not met. The WWC analysis found no statistically significant negative effects in this domain.

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

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