

Appendix

Appendix A1 Study characteristics: Demetriades-Guyette, 2002 (quasi-experimental design)

Characteristic	Description
Study citation	Demetriades-Guyette, A. (2002). Patterns of change in the social-cognitive development of middle school children following a school-based multicultural literature program. <i>Dissertation Abstracts International</i> , 63(05B), 2615. (UMI No. 3052695)
Participants	The study included 100 sixth- and seventh-grade students from five middle schools. About 50% of the sample was female. A higher percentage of minority students were in the intervention group (77%) than in the comparison group (57%). The largest minority group in both the intervention (35%) and comparison conditions (20%) defined themselves as multiracial. The second largest minority group was African-American students in the intervention group (14%) and Asian students in the comparison group (13%).
Setting	Both the intervention and comparison schools were part of Cambridge Public Schools, Massachusetts.
Intervention	The program consisted of a 12-week literature-based curriculum. The program was co-taught by the regular classroom teachers and a staff developer from the Cambridge Youth Guidance Center.
Comparison	The comparison schools were drawn from the same school districts as the intervention schools and were matched on demographic characteristics. Comparison schools did not implement the <i>Voices Literature and Character Education Program</i> but intended to implement the program after the end of the study.
Primary outcomes and measurement	The study investigated student outcomes in the knowledge, attitudes, and values domain using the GSID Relationship Questionnaire (REL-Q). This student survey included five subscales measuring interpersonal understanding and interpersonal skills. (See Appendix A2 for more detailed descriptions of the outcome measures.) ¹
Teacher training	Teachers participated in a three-day training prior to program implementation.

1. The study also used a behavior count questionnaire assessing frequencies of negative and prosocial behaviors. The study author reported no statistically significant effects for these student outcomes. The WWC could not review these measures because means and standard deviations were not available for review.

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

Outcome measure	Description
The GSID Relationship Questionnaire: maturity (best response score) scale	GSID Relationship Questionnaire (as cited in Demetriades-Guyette, 2002): Relationship Maturity, scored by “best response.” Total score on a 24-item survey (Rel-Q) developed by the Group for the Study of Interpersonal Development (GSID). The survey consists of five subscales. Two subscales represent interpersonal understanding (understanding and perspective-taking), two subscales represent interpersonal skills (hypothetical and real-life interpersonal negotiation), and the fifth Rel-Q subscale represents personal meaning awareness. The best response score is based on the student’s choice of the best response of four possible responses to each question.
The GSID Relationship Questionnaire: maturity (item rating score) scale	GSID Relationship Questionnaire (as cited in Demetriades-Guyette, 2002): Relationship Maturity, scored by “item rating.” Total score on a 24-item survey (Rel-Q) developed by the Group for the Study of Interpersonal Development (GSID). The survey consists of five subscales. Two subscales represent interpersonal understanding (understanding and perspective-taking), two subscales represent interpersonal skills (hypothetical and real-life interpersonal negotiation), and the fifth Rel-Q subscale represents personal meaning awareness. The response rating score is based on the student’s assignment of “poor,” “average,” “good,” and “excellent” to each of four responses to each question.

Appendix A3 Summary of study findings included in the rating for the knowledge, attitudes, and values domain¹

Outcome measure	Study sample	Sample size (students/schools)	Author's findings from the study					
			Mean outcome (Standard deviation ²)		WWC calculations			
			<i>Voices LACE</i> group ³ (column 1)	Comparison group (column 2)	Mean difference ⁴ (column 1–column 2)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
Demetriades-Guyette, 2002 (quasi-experimental design)								
Relationship maturity (best response score)	Grades 6–7	98 students/ 5 schools	2.10 (0.31)	2.05 (0.31)	0.05	0.16	ns	+6
Relationship maturity (item rating score)	Grades 6–7	98 students/ 5 schools	2.09 (0.18)	2.06 (0.15)	0.03	0.16	ns	+6
Domain average⁸ for knowledge, attitudes, and values						0.16	ns	+6

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index. Findings on the individual item level are presented in Appendix A4.
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.
3. The WWC requested and received means and standard deviations for all outcomes in the knowledge, attitudes, and values domain, because they were not reported in the paper.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
6. 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 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 statistical significance. In the case of the *Voices* report, no corrections for clustering or multiple comparisons were needed.
7. 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.
8. 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 for the knowledge, attitudes, and values domain¹

Outcome measure	Study sample	Sample size (students/schools)	Author's findings from the study					Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			Mean outcome (Standard deviation ²)		WWC calculations				
			Voices LACE group ³ (column 1)	Comparison group (column 2)	Mean difference ⁴ (column 1–column 2)	Effect size ⁵			
Demetriades-Guyette, 2002 (quasi-experimental design)									
Interpersonal understanding (best choice score)	Grades 6–7	98 students/ 6 schools	1.95 (0.43)	1.91 (0.40)	0.04	0.09	ns	0.04	
Hypothetical negotiation (best response score)	Grades 6–7	98 students/ 6 schools	2.00 (0.58)	2.04 (0.07)	–0.04	–0.10	ns	–0.04	
Real life negotiation (best response score)	Grades 6–7	98 students/ 6 schools	2.13 (0.60)	2.16 (0.64)	–0.03	–0.05	ns	–0.02	
Personal meaning (best response score)	Grades 6–7	98 students/ 6 schools	2.12 (0.41)	1.95 (0.57)	0.17	0.34	ns	0.13	
Perspective coordination (best response score)	Grades 6–7	98 students/ 6 schools	2.22 (0.49)	2.05 (0.45)	0.17	0.36	ns	0.14	
Interpersonal understanding (item rating score)	Grades 6–7	98 students/ 6 schools	2.03 (0.22)	2.04 (0.19)	0.00	–0.02	ns	–0.01	
Hypothetical negotiation (item rating score)	Grades 6–7	98 students/ 6 schools	2.15 (0.28)	2.14 (0.24)	0.01	0.02	ns	0.01	
Real life negotiation (item rating score)	Grades 6–7	98 students/ 6 schools	2.13 (0.25)	2.16 (0.64)	–0.03	–0.06	ns	–0.02	
Personal meaning (item rating score)	Grades 6–7	98 students/ 6 schools	2.01 (0.23)	1.97 (0.18)	0.03	0.16	ns	0.07	
Perspective coordination (item rating score)	Grades 6–7	98 students/ 6 schools	2.17 (0.26)	2.10 (0.21)	0.07	0.30	ns	0.12	

ns = not statistically significant

1. This appendix presents item-level findings for measures that fall in the knowledge, attitudes, and values domain. Aggregated scale scores were used for rating purposes and are presented in Appendix A3.
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.
3. The WWC requested and received means and standard deviations for all outcomes in the knowledge, attitudes, and values domain, because they were not reported in the paper.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
6. 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 corrects for clustering within classrooms or schools. 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 statistical significance. In the case of the *Voices* report, no corrections for clustering were needed.
7. 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 *Voices Literature and Character Education Program* 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 the single study that met WWC standards did not show 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. *Voices Literature and Character Education Program* had only one evaluation study meeting WWC evidence standards with reservations that reported findings on knowledge, attitudes, and values.

- 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. The WWC analysis found no statistically significant or substantively important negative effects in this domain. Because one study showed indeterminate effects and no studies showed statistically significant or substantively important positive effects, *Voices Literature and Character Education Program* did not meet this criterion.

(continued)

Appendix A5 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 one study that met WWC standards with reservations did not show 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 one study that met WWC standards with reservations did not show 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 one study that met WWC standards with reservations did not show 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.

Not met. The one study that met WWC standards with reservations did not show statistically significant or substantively important negative 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 one study that met WWC standards with reservations did not show statistically significant negative effects in this domain.

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

Met. The one study that met WWC standards with reservations did not show statistically significant or substantively important positive 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.