

Table S1
Study Aim(s)

Authors (Year)	Research Question(s) or Purpose(s)
Castro-Olivo and Merrell (2012)	1) Assess the program's feasibility of implementation in a regular classroom setting 2) Assess the program's effectiveness at teaching SEL skills 3) Assess the program's social validity and participant satisfaction
Castro-Olivo (2014)	1) To what extent do students who participate in the culturally adapted Jóvenes Fuertes SEL program report higher levels of SEL knowledge and SEL resiliency in comparison to those in the waitlist control group at post intervention assessments? 2) To what extent do participants perceive the culturally adapted Jóvenes Fuertes SEL program as a socially valid intervention
Castro-Olivo et al. (2022)	1) To what extent the proposed adapted factor structure of the BERS-2 adequately measures each of the 5 SEL competencies (i.e. relationship skills, self-awareness, responsible decision making, social awareness, and self-management) in this sample? 2) To what extent does a culturally-adapted SEL intervention improve each SEL competency (i.e. relationship skills, self-awareness, responsible decision making, social awareness, and self-management) for this ELL sample?
Cramer and Castro-Olivo (2015)	1) Is there a significant difference between pre-intervention and post-intervention self-ratings of student resiliency? 2) Does a significant difference exist between post-intervention and follow-up ratings of student resiliency? 3) Do students report a significant difference between pre-intervention and post-intervention ratings of internalizing problem symptoms? 4) Does a significant difference exist between post-intervention and follow-up ratings of student internalizing problem symptoms? 5) To what extent do CLD students find the culturally adapted Strong Teens curriculum socially valid?
Elek et al. (2010)	This study examined the effects of an enhanced version of the Keepin' it REAL curriculum to addresses issues related to the transition faced by youth experiencing acculturation
Graves et al. (2017)	1) Does implementation of Strong Start result in an increase in knowledge of social and emotional skills in urban African American students?

Authors (Year)	Research Question(s) or Purpose(s)
	2) Does Strong Start increase the social-emotional assets in urban African American students? 3) Does implementing Strong Start decrease externalizing behavior in urban African American students?
Hecht et al. (2018)	The study examined the relationship between delivery quality and program adaptation.
Hopson (2008)	1) Is an adapted version of Keepin' it REAL able to reduce intentions to accept offers of substances among participating alternative school students? 2) Is an adapted version of Keepin' it REAL able to reduce substance use among participating alternative school students?
Kurtz (2020)	1) What are the opinions of involved parties of the adapted Strong Start and Strong Kids lesson plans for implementation in an urban school? 2) Are there significant differences in outcomes for students who receive the adapted Strong Start and Strong Kids intervention class wide versus in a small group, as measured by Negative Affect scores on the universal behavioral health screener: Behavior Intervention Monitoring Assessment System 2 (BIMAS-2)? 3) Are there significant differences in outcomes for students who receive the adapted Strong Start and Strong Kids intervention versus the district mean, as measured by Negative Affect scores 4) Do changes in Negative Affect scores on the BIMAS-2 in response to the adapted intervention maintain from post-intervention to follow up?
Martinez (2011)	This study was a program evaluation that appraised the Spark Program for its potential to produce significant changes in at-risk children's emotional and behavioral outputs in key areas such as anger management, educational aspirations, peer attachment and self-esteem.
Moran et al. (2019)	1) Do students trained in the adapted STAC intervention report an increase in knowledge and confidence to intervene as "defenders?" 2) Do students trained in the culturally adapted STAC intervention use the STAC strategies when they observe bullying? 3) Were there differences in outcomes between White and Hispanic students? 4) What were students' experiences participating in the culturally adapted STAC intervention and using the STAC strategies to intervene in bullying situations?

Note. The author(s)' language was used when possible for the research questions and purposes column. The authors of this systematic review recognize that language is ever-changing. There is currently controversy around potential colonialist roots to the term "stakeholder" ([Reed, 2022](#)). Thus, the term "stakeholder" from the Kurtz (2020) dissertation was intentionally changed to "involved parties."

Table S2

Study Setting and Design

Authors (Year)	Setting of school(s) and/or district(s)	Design	
		Analysis Type	Method
Castro-Olivo and Merrell (2012)	Schools (N = 1) East Bay area of Northern California	Within group	Pre-post intervention design
Castro-Olivo (2014)	Schools and Districts (N = 7 schools in N = 3 Districts) Southern California	Between group: intervention vs control	Quasi-experimental intervention/control pre-post
Castro-Olivo et al. (2022)	Schools and Districts (N = 7 schools in N = 3 Districts) Southern California	Between group: intervention vs control group	Secondary data analysis (original study was quasi-experimental)
Cramer and Castro-Olivo (2015)	Schools (N = 1) Suburban, Southern California	Within group	Quasi-experimental pre- and post-intervention group design
Elek et al. (2010)	Schools (N = 30) Phoenix, Arizona	Between group: 5 th grade implementation vs control/ 7 th grade implementation only vs control/ 5 th and 7 th grade implementation vs control	Quasi-experimental (longitudinal)
Graves et al. (2017)	Schools (N = 1) Urban, Mid-Atlantic Region	Between group: intervention vs control	Randomized delayed treatment control

Authors (Year)	Setting of school(s) and/or district(s)	Design	
		Analysis Type	Method
Hecht et al. (2018)	Schools (N = 39) Rural Pennsylvania and Ohio	Between group: non-adapted curriculum vs adapted curriculum vs control	Randomized control trial
Hopson (2008)	Schools (N = 4) Urban (<i>n</i> = 1 school) Other schools not specified	Between group: intervention vs control	Mixed methods, quasi-experimental pretest posttest design with a six-week follow-up and qualitative procedures
	Schools (N = 1) Urban, Boston, Massachusetts	Between group: small group vs class wide - Small group: one combined second and third grade separate classroom for students with learning disabilities and Autism Spectrum Disorder, one combined fourth and fifth grade separate classroom for students with learning disabilities and Autism Spectrum Disorder, one general education fifth grade classroom, one second grade inclusion classroom - Classwide: one kindergarten inclusion classroom, one third grade Sheltered English Immersion classroom, one fourth grade general education classroom	Quasi-experimental group design
Kurtz (2020)			
Martinez (2011)	Schools (N =2) Bay Area, California	Within group	Secondary analysis

Authors (Year)	Setting of school(s) and/or district(s)	Design	
		Analysis Type	Method
Moran et al. (2019)	Schools (N = 1) Rural, Northwestern, U.S.	Within group	Mixed methods

Table S3
Participant Information

Author (Year)	Total	Grade	Race and/or Ethnicity	Gender or Sex Assigned at Birth
Castro-Olivo and Merrell (2012)	N = 40	High school	Latino, recent immigrant (100%)	Female (50%) Male (50%)
Castro-Olivo (2014)	N = 102 Intervention: <i>n</i> = 49 Control: <i>n</i> = 53	Middle school and high school	Latino (100%); Intervention: Mexican origin (15%) United States Origin (15%) Other origin (3%) Control: Mexican origin (25%) United States origin (25%) Other origin (1%)	Intervention: Boys (26%) Girls (20%) Did not report (0%) Control: Boys (23%) Girls (29%) Did not report (1%)
Castro-Olivo et al. (2022)	N = 102 Intervention: <i>n</i> = 49 Control: <i>n</i> = 53	Middle school and high school	Latino (100%) Intervention: Mexican origin (15%) United States Origin (15%) Other origin (3%) Control: Mexican origin (25%) United States origin (25%) Other origin (1%)	Intervention: Boys (26%) Girls (20%) Did not report (0%) Control: Boys (23%) Girls (29%) Did not report (1%)
Cramer and Castro-Olivo (2015)	N = 20	9 th and 10 th grade	Latino/Hispanic (75%) African American (15%) Caucasian (5%) No response (5%)	Female (20%) Male (20%)

Author (Year)	Total	Grade	Race and/or Ethnicity	Gender or Sex Assigned at Birth
			United States origin (75%) Mexican origin (25%)	
Elek et al. (2010)	N = 1,984	5 th grade & 7 th grade	Mexican or Mexican American (75%) Other Hispanic background (3%) non-Hispanic white (5%) African American (9%) Native American (3%)	Female (50%) Male (50%)
Graves et al. (2017)	N = 61	Kindergarten, 1 st , and 2 nd grade	African American (100%)	Male (100%)
Hecht et al. (2018)	N = 2781 Non-adapted: <i>n</i> = 664 Low quality: <i>n</i> = 329 High quality: <i>n</i> = 335 Adapted: <i>n</i> = 1022 Low quality: <i>n</i> = 590 High quality: <i>n</i> = 432 Control: <i>n</i> = 1095	7 th , 8 th , and 9 th grade	European American (97%)	Female (49%) Male (51%)
Hopson (2008)	N = 41 Intervention: <i>n</i> = 18 Control: <i>n</i> = 23	High school	Intervention: Black/African American (5.6%) Hispanic (38.9%) white (38.9%) other (16.7%) Control: Black/African American (8.7%) Hispanic (47.8%) white (30.4%)	Intervention: female (61.1%) male (38.9%) Control : female (56.5%) male (43.5%)

Author (Year)	Total	Grade	Race and/or Ethnicity	Gender or Sex Assigned at Birth
			other (13%)	
Kurtz (2020)	N = 43	Kindergarten, 2 nd , 3 rd , 4 th , 5 th grade	Hispanic (81.4%) white (18.6%)	Female (51.16%) Male (48.84%)
Martinez (2011)	N = 38	4 th grade – 8 th grade	Latino/Hispanic (68%) African-American/Black (13%) white/Caucasian (2%) Other (13%)	Female (36.8%) Male (63.2%)
Moran et al. (2019)	N = 63 Qualitative: <i>n</i> = 23 (subset of the N = 63)	Elementary school	Quantitative analysis: white (55.4%) Hispanic (44.6%) Qualitative analysis: white (65%) Hispanic (35%)	Quantitative analysis females (57.1%) males (41.3%) other (1.6%) Qualitative analysis: female (65%) male (35%)

Note. In terms of race/ethnicity and gender or sex assigned at birth, the original author language was kept. However, given the gender-inclusivity of the term “Latine”, the Martinez (2011), Castro-Olivo and Merrel (2012), Castro-Olivo (2014), Cramer and Castro-Olivo (2015), and Castro-Olivo et al. (2022) had the term “Latino” replaced with “Latine.” The authors of this systematic review also recognize there is still much controversy around terms to use for this population ([López, 2022](#); [Yarin, 2022](#))

Table S4
Implementer Characteristics

Study	Total	Role	Years Experience in Role	Experience with Implemented SEL program	Race/Ethnicity	Gender or Sex Assigned at Birth
Castro-Olivo and Merrell (2012)	N = 2	Teachers	Missing	“Minimal training”	Missing	Missing
Castro-Olivo (2014)	N = 2	Teachers	Missing	4-hour training	Latina (100%)	Women (100%)
Castro-Olivo et al. (2022)	N = 2	Teachers	Missing	4-hour training	Latina (100%)	Women (100%)
Cramer and Castro-Olivo (2015)	N = 2	Graduate students	Missing	4-hour training	Missing	Missing
Elek et al. (2010)	Missing	Missing	Missing	Missing	Missing	Missing
Graves et al. (2017)	N = 3	Doctoral graduate students (n = 2) Professor/graduate student (n = 1)	Missing	3-day training	Missing	Missing
Hecht et al. (2018)	N = 32	Teachers	Missing	1 day training	Missing	Missing
Hopson (2008)	N = 6	Teachers or counselors	Missing	60–90-minute training	Missing	Missing

Study	Total	Role	Years Experience in Role	Experience with Implemented SEL program	Race/Ethnicity	Gender or Sex Assigned at Birth
Kurtz (2020)	N = 10	Teachers (n = 7) Researchers (n = 3)	Teachers: 1 year (n = 1) 5 years (n = 1) 6 years (n = 2) 7 years (n = 1) 11 years (n = 1) 19 years (n = 1) Researchers: Missing	No experience	Black/African American/Hispanic (14%) White (86%)	Missing
Martinez (2011)	N = 8	Clinical psychology doctoral students	Missing	Missing	Missing	Missing
Moran et al. (2019)	Missing	Graduate students	Missing	Missing	Missing	Missing

Note. MISSING: The information is not mentioned in the article at all; NOT EXAMINED: The authors explicitly mention they did not analyze a variable; NOT REPORTED: The authors had a measure, but did not report findings; NOT REPORTED AS EXPECTED: The way the data were reported were not in the same unit of study as other variables (e.g. reports demographic for the entire school instead of for the group)

Table S5
Quality Appraisal

Authors (Year)	A	B	C	D	E	F	G	H	I	J	K	L	M	Total
Castro-Olivo and Merrell (2012)	1	3	3	3	2	3	3	3	1	0	3	3	3	31
Castro-Olivo (2014)	3	3	2	3	2	3	3	2	1	0	3	3	3	31
Castro-Olivo et al. (2022)	3	3	3	3	3	3	3	3	3	3	3	3	2	38
Cramer and Castro-Olivo (2015)	3	3	2	3	2	3	3	3	2	3	3	3	3	36
Elek et al. (2010)	1	2	3	3	1	2	2	1	0	0	3	3	3	24
Graves et al. (2017)	1	3	2	3	3	3	3	2	3	0	3	3	2	31
Hecht et al. (2018)	3	3	2	3	3	1	3	3	3	3	3	3	3	36
Hopson (2008)	3	3	2	3	3	3	3	3	3	3	3	3	3	38
Kurtz (2020)	2	3	3	3	3	3	3	3	3	3	3	3	3	38
Martinez (2011)	3	3	2	3	2	3	3	3	3	0	3	3	3	34
Moran et al. (2019)	1	3	2	3	3	3	3	3	3	3	3	3	3	36

Note. Per item low score = 0; Per item high score = 3; Max score = 39

A = Theoretical or conceptual underpinning to the research; **B** = Statement of research aim(s); **C** = Clear description of research setting and target population; **D** = The study design is appropriate to address the stated research aim(s); **E** = Appropriate sampling to address the research aim(s); **F** = Rationale for choice of data collection tool(s); **G** = The format and content of data collection tool is appropriate to address the stated research aim(s); **H** = Description of data collection procedure; **I** = Recruitment data provided; **J** = Justification for analytic method selected; **K** = The method of analysis was appropriate to answer the research aim(s); **L** = Evidence that those involved or

impacted by the research have been considered in research design or conduct; **M** = Strengths and limitations critically discussed; **Total** = Total score on the QUADS

Table S6

SEL Program Descriptives

Authors (Year)	SEL Program/ Adapted Name (if applicable)	Program Type (as used in the study)	Number of Lessons	Length of Program	Targeted CASEL competencies/other competencies
Castro-Olivo and Merrell (2012)	Strong Teens/Jóvenes Fuertes	Class-wide	12 lessons	12 weeks◆	Self-management* (e.g. stress reduction) Self-awareness* (e.g. emotional awareness) Responsible decision making* (e.g. problem solving) Relationship skills* (e.g. problem solving) Social awareness* (e.g. empathy training)
Castro-Olivo (2014)	Strong Teens/Jóvenes Fuertes	Class-wide	12 lessons	12 weeks	Self-management* (e.g. dealing with anger) Self-awareness* (e.g. clear thinking) Responsible decision making* (e.g. problem solving) Relationship skills* (e.g. problem solving) Social awareness* (e.g. understanding other's emotions)
Castro-Olivo et al. (2022)	Strong Teens/Jóvenes Fuertes	Class-wide	12 lessons	12 weeks	Self-management Self-awareness Responsible decision making Relationship skills Social awareness
Cramer and Castro-Olivo (2015)	Strong Teens	Small group	12 lessons	6 weeks	Self-management Self-awareness Responsible decision making Relationship skills Social awareness

Authors (Year)	SEL Program/ Adapted Name (if applicable)	Program Type (as used in the study)	Number of Lessons	Length of Program	Targeted CASEL competencies/other competencies
Elek et al. (2010)	Keepin' it REAL/keepin' it REAL- acculturation enhanced	School- wide	12 lessons	Not specified (45 mins/les son)	Self-management* (e.g. refusal self-efficacy) Self-awareness* (e.g. learning normative beliefs) Responsible decision making* (e.g. model resistance strategies) Social awareness* (e.g. learning normative beliefs) /anti-drug expectancies
Graves et al. (2017)	Strong Start	Small group	14 lessons	14 weeks	Self-management* (e.g. handling uncomfortable emotions) Self-awareness* (e.g. identify body, facial, and situational cues) Responsible decision making* (e.g. solving problems with friends) Relationship skills* (e.g. making and keeping friends) Social awareness* (e.g. labeling others emotions)
Hecht et al. (2018)	Keepin' it REAL/keepin' It REAL – Rural Adaptation	Class-wide	Missing	Missing	Responsible decision-making* (e.g. preventing substance use)
Hopson (2008)	Keepin' it REAL	Class-wide	6 lessons	6 weeks	Self-management* (e.g. refusal strategies) Self-awareness* (e.g. explain reasons for refusing) Responsible decision making* (e.g. avoid substance situations) Relationship skills* (e.g. explain reasons for refusing) Social awareness* (e.g. avoiding substance use situations)
Kurtz (2020)	Strong Start and Strong Kids	Class-wide and small group	12 lessons	12 weeks	Self-management* (e.g. coping with stress) Self-awareness* (e.g. understanding personal emotions) Responsible decision making* (e.g. clear thinking) Relationship skills* (e.g. problem solving) Social awareness* (e.g. understanding others emotions)
Martinez (2011)	Spark Program	Small group	Not specified,	1 school year	Self-management* (e.g. setting high, reasonable goals) Self-awareness* (e.g. anger management)

Authors (Year)	SEL Program/ Adapted Name (if applicable)	Program Type (as used in the study)	Number of Lessons	Length of Program	Targeted CASEL competencies/other competencies
			weekly, 45-60 minute lessons		Responsible decision making* (e.g. resisting peer pressure) Relationship skills* (e.g. building trust) Social awareness* (e.g. using communication skills)
Moran et al. (2019)	STAC (Stealing the show, Turning it over, Accompanying others, and Coaching compassion)	Small group	3 lessons	2 hours	Self-management* (e.g. using humor to interrupt bullying) Self-awareness* (e.g. understanding bystander roles) Responsible decision making* (e.g. getting help from adult) Relationship skills* (e.g. offering support to target of bullying) Social awareness* (e.g. knowing why students bully)

Note. ♦ indicates information received by emailing the first author. * Indicates implicit discussion of competency. Some examples may be repeated as several CASEL competencies were suitable for the skills these programs taught

Table S7
Study Treatment Integrity

Authors (Year)	Treatment Integrity	
	Measure	Outcome
Castro-Olivo and Merrell (2012)	12-item observation tool completed by the first-author over 25% of lessons	100% (related to the 12-item observation)
	Audiotapes of sessions—teachers determined amount of time spent on each concept	Teachers spent recommended amount of time for all main concepts; measured by (related to the audiotapes of sessions)
	Teacher survey	Teachers felt they were “good” to “very good” at managing time to achieve lesson agendas and skill development, and teachers felt their overall effectiveness at teaching the program was “good to very good” (related to the teacher survey)
Castro-Olivo (2014)	Checklist of all main goals for each lesson, completed by principal investigator	Not reported
Castro-Olivo et al. (2022)	Authors state the original study (see: Castro-Olivo (2014)) measured fidelity for each lesson.	Not reported
Cramer and Castro-Olivo (2015)	Missing	Missing
Elek et al. (2010)	Observers (no further detail was provided)	Qualitative report only: In general, observers rated implementation quality uniformly high. Teachers did not provide consistent student attendance reports.
Graves et al. (2017)	Unspecified means of measurement: First author completed fidelity checks for approximately 20% of all randomly selected sessions	100%
Hecht et al. (2018)	Missing	Missing

Authors (Year)	Treatment Integrity	
	Measure	Outcome
Hopson (2008)	The researcher conducted several, brief, 15-minute observations, and implementer interviews were conducted in person or via email	Qualitative report only: In general, most schools implemented with fidelity. One school did not do activities as student interaction was not allowed; two schools changed a few scenarios to make them more relatable; one school did not finish implementing the curriculum
Kurtz (2020)	Measured by number of partially and fully implemented steps divided by total number of applicable intervention steps for each session and multiplied by 100 for each group/teacher	Overall fidelity ($M = 97.14\%$, $SD = 2.59$)
Martinez (2011)	Not reported	Not reported
Moran et al. (2019)	Missing	Missing

Note. MISSING: The information is not mentioned in the article at all; NOT EXAMINED: The authors explicitly mention they did not analyze a variable; NOT REPORTED: The authors had a measure, but did not report findings; NOT REPORTED AS EXPECTED: The way the data were reported were not in the same unit of study as other variables (e.g. reports demographic for the entire school instead of for the group)

Table S8*Adaptations and Adaptation Purposes*

Authors (Year)	Adaptation Type(s)	Adaptation Purpose(s)
Castro-Olivo and Merrell (2012)	Surface: Bilingual implementer Surface: Materials translated to Spanish Surface: Bicultural implementer Deep: Incorporating cultural values in SEL program design -All examples adapted to reflect cultural values, costumes, and traditions Deep: Incorporating cultural values in SEL program implementation -Teaching program's target skills altered to reflect cultural reframing often used by Latine populations (e.g. respeto and familismo) Deep: Other -Included metaphors and folktales -Taught concepts of acculturative stress and ethnic pride (brings in cultural reality)	Created for a more appropriate use with a Latine Immigrant population
Castro-Olivo (2014)	Surface: Bilingual implementer Surface: Intervention and materials translated into Spanish Surface: Bicultural implementer Deep: Incorporating cultural values in SEL program design -All examples were adapted to reflect cultural values, costumes, and traditions Deep: Incorporating cultural values in SEL program implementation -Teaching program's target skills altered to reflect the cultural reframing often used by Latine populations (e.g. respeto and familismo) Deep: Other	Wanted to validate the Jóvenes Fuertes with a Latine ELL sample enrolled in middle and high school students

Authors (Year)	Adaptation Type(s)	Adaptation Purpose(s)
	-Included metaphors and folktales -Taught concepts of acculturative stress and ethnic pride (brings in cultural reality)	
Castro-Olivo et al. (2022)	Surface: Bilingual implementer Surface: Intervention and materials translated into Spanish Surface: Bicultural implementer Surface: Deep: Incorporating cultural values in SEL program design -All examples adapted to reflect cultural values, costumes, and traditions Deep: Incorporating cultural values in SEL program implementation -Teaching program's target skills altered to reflect the cultural reframing often used by Latine populations (e.g. respeto and familismo) Deep: Other -Included metaphors and folktales -Taught concepts of acculturative stress and ethnic pride (brings in cultural reality)	Aim was to validate the Jóvenes Fuertes with a Latino ELL sample enrolled in middle and high school students
Cramer and Castro-Olivo (2015)	Surface: Other -Used culturally appropriate language (choice between English and Spanish) -Implementers trained in identifying and being sensitive to needs of the group Deep: Incorporating cultural values in SEL program design -Used cultural knowledge to align intervention procedures to increase acceptability Deep: Incorporating cultural values in SEL program implementation	Conducted the cultural adaptations for a predominantly diverse group of culturally and linguistically diverse (CLD) students who were no longer identified/classified as English language learners (ELL's)

Authors (Year)	Adaptation Type(s)	Adaptation Purpose(s)
	-Encouraged students to consider their own cultural values, customs, and traditions -Considered cultural values that relate to goals for home and school -Considered culture-specific life and family circumstances (e.g. immigration status) Deep: Other -Used cultural metaphors of the target group and explain use if not understood by other groups -Taught concepts of acculturative stress and ethnic pride (brings in cultural reality)	
Elek et al. (2010)	Deep: Incorporating cultural values in SEL program design -Promoted relevant protective cultural values (such as family centeredness or familismo to reflect the study schools predominantly Mexican and Mexican American populations) Deep: Other -Added two additional lessons centering acculturation -Emphasized cultural diversity and ethnic identity as strengths -Examined the impact of language on drug resistance and discussions with parents -Explored the changes in identity and values that may occur through acculturation	Addressed the specific needs of students experiences acculturation
Graves et al. (2017)	Surface: Other -Implemented at the Tier II instead of Tier 1 level Deep: Incorporating cultural values in SEL program implementation -Made literature modifications	Enhanced the curriculum relevance for African American sample Skills taught in the intervention aligned well with teacher desired skills for students to learn

Authors (Year)	Adaptation Type(s)	Adaptation Purpose(s)
Hecht et al. (2018)	Deep: Incorporating cultural values in SEL program design -Collaborated with rural teachers to integrate rural culture surrounding drug use into the curriculum (including role play activities, decision scenarios, and homework, curriculum design, and intervention strategy)	Aim was to integrate information into the program on how rural cultures differ from urban and suburban counterparts (ie. health disparities, weaker protective factors)
Hopson (2008)	Surface: Other -Adapted teacher version of the manual to make 6 sessions Deep: Incorporating cultural values in SEL program design -Held focus groups and group activities to describe personal stories -Student lead scenario development -Recreated videos that fit student experiences (recreated by students) -Created new role play scenarios (led by students)	Collapsed the curriculum into a six-week session as alternative school sessions typically only go for 6-8 weeks, so a 10-week intervention was not feasible or culturally appropriate Created an adapted program to be culturally grounded for youth in Texas Made the curriculum appropriate for older age students (14-19 year olds) in alternative schools
Kurtz (2020)	Surface: Translation of materials -Provided the book "The Way I Feel" in Spanish -Family bulletin translated into Spanish to send home with students Surface: Inclusion of ethnic lifestyle elements -Asked students to share synonyms for angry in the language they use at home Surface: Inclusion of same race/ethnicity role model -Chose an alternative book where the main character is a child of color: Marvelous Me by Lisa Bullard -Chose an alternative book the main character is a child of color: Those Shoes by Maribeth Boelts Surface: Other -Used individual body maps to identify how feelings are experienced in the body	Focus group was implemented to guide the intervention adaptation process Focused on internalizing problems in the context of a high needs urban elementary school

Authors (Year)	Adaptation Type(s)	Adaptation Purpose(s)
	-Included additional mindfulness activity and coping skill instruction (e.g., breathing, squeezing clay) Deep: Incorporating cultural values in SEL program design -Got community feedback on researcher created adaptations -Included culturally relevant examples for having multiple feelings at once (e.g., when someone has to leave Boston) -Included psychoeducation about how experiencing trauma can make it difficult to identify feelings when all feelings can feel like angry or scared Deep: adjusting materials to literacy level of participant -Incorporated extension activity “Mirror Mirror” to support English Language Learners who may require more visuals Deep: Other -Asked students to share physical clues related to how people at home show different emotions -Taught the trauma response of “flight, fight, or freeze” -Discussed the link between trauma and anger, and how trauma can make us feel like many situations/experiences are “big deals” despite them being “little deals” -Included psychoeducation about how experiencing trauma can make it difficult to identify physical clues of emotions in others	
Martinez (2011)	Surface: other -Lengthened course of treatment (to support community with high need) -Multiple modes of treatment options (started group, gave option for individual support) Deep: Involving family in SEL program -Formal graduation party for students and their families	Adapted to fit the specific needs of the diverse, at-risk populations of children and youth who attended two local public schools (one elementary school and one junior high school) in the Bay Area Extended treatment: schools had limited resources and high need

Authors (Year)	Adaptation Type(s)	Adaptation Purpose(s)
	Deep: Other -Advocacy efforts (e.g. involved in IEP meetings, translation services)	
Moran et al. (2019)	Surface: Bilingual implementer -Recruited ethnically diverse and bilingual graduate students who spoke English and Spanish Surface: Inclusion of ethnic lifestyle elements -Infusion of culturally relevant language (e.g. used “defender” instead of “advocate”) Surface: Other -Trained delivery consistent with cultural norms and values -Trainers encouraged to be intentional about fostering connections with and between students to promote relationship building Deep: Incorporating cultural values in SEL program design -Inclusion of culturally relevant bullying experiences and training goals -Included examples related to culture and ethnicity	In response to the growing need for relevant and accessible interventions in ethnically blended, rural, low-income schools, the researchers adapted the STAC program to be culturally appropriate for this population

Note. The adaptation purposes were extracted to retain as much of the studies’ original verbiage as possible. Bolded headings are the common features of surface and deep level features. Listed below are extracted examples from the reviewed studies.

Table 9
Student Outcomes

Authors (Year)	Dependent Variables	Measures	Outcomes of Measures (Quantitative, if possible, otherwise qualitative provided)	Author Description of Findings
Castro-Olivo and Merrell (2012)	-Knowledge of the SEL content taught in program -Participant internalizing symptoms -Acculturative stress -Students' sense of belonging/connec tedness to school	- Internalizing symptoms test -Strong Teens knowledge test -Societal attitude familial environments-for children (SAFE-C; Spanish version) -People in my life scale (sense of school belonging subscale)	- Knowledge of SEL content taught in the program: ($d = 0.98$, large effect)*** - Internalizing disorders : ($d = 0.15$) - Acculturative stress : ($d = 0.20$, small effect) - School belonging : ($d = 0.37$, small effect)	- Knowledge of SEL : Statistically significant SEL knowledge was gained from baseline to post- intervention. - Internalizing disorders : No statistical significance was found for students' reported levels of symptoms of internalizing disorders. - Acculturative stress : No statistical significance was found for students' reported levels of symptoms of internalizing disorders. - School belonging : No statistical significance was found for students' reported levels of symptoms of sense of school belonging.
Castro-Olivo (2014)	-Social emotional resiliency -Student knowledge of SEL concepts	-BERS-2 -Strong Teens Knowledge Test— Spanish version	- Social emotional resiliency : ($d = .44$)* - Student knowledge of SEL concepts ($d = 1.27$)*	Students assigned to the intervention condition demonstrated significantly more resilience and knowledge.
Castro-Olivo et al. (2022)	-Self-awareness -Relationship skills -Self-management -Social awareness	- BERS-2 (Behavioral and emotional rating scale) Spanish version,	Intervention vs control group: ($d =$ pre intervention/post intervention) - ▲ Self-awareness ($d = .23/.662$)* - ▲ Relationship skills ($d = -0.0017/0.39$)** - ▲ Self-management ($d = 0.187/0.0725$)	Students in the intervention condition showed significantly higher-levels of self-awareness and relationship skills.

Authors (Year)	Dependent Variables	Measures	Outcomes of Measures (Quantitative, if possible, otherwise qualitative provided)	Author Description of Findings
	-Responsible decision making	restructured for this study	-▲Social awareness ($d = 0.1799/0.2276$) -▲Responsible decision making ($d = 0.0638/.1043$)	
Cramer and Castro-Olivo (2015)	-Student resiliency -Internalizing symptoms	- BERS-2 YRS (Behavioral and emotional rating scale, youth rating scale) -CBCL, youth self-report only (24 items only, not full scale)	- Student resiliency: Before and after intervention scores: $t(19) = -2.12, p = 0.048, ES = 0.2$; post-intervention and follow-up scores: $t(19) = -0.23, p = 0.822$ - Internalizing symptoms: Before and after intervention scores: $t(19) = 0.82, p = 0.424$; post-intervention and follow up scores: $t(19) = 1.17, p = 0.256$	- Student resiliency: There was a significant difference between resiliency scores before and after the intervention. There was no significance between post-intervention resiliency scores and follow-up resiliency scores was not significant. - Internalizing symptoms: There was not a statistically significant difference between before and after intervention scores. There was also not a statistically significant difference between post-intervention and follow-up internalizing symptom scores.
Elek et al. (2010)	- Lifetime substance use prevalence -Past month's prevalence -Intentions to use substances -Refusal efficacy -Hypothetical alcohol resistance -Number of substance use	- Long scannable questionnaire (104-177 items) with items adapted from Kasen, Vaughan, and Walter (1992), Hecht et al. (2003), and Hansen and Graham (1991)	Effect sizes (5th grade only vs control/ 7th grade only vs control/ 5th+7th vs control) - Past month's substance use prevalence ($d = .48^*/-.23/-0.05$) - Lifetime substance use prevalence ($d = .11/-0.20/-0.18$) - Intentions to use substances ($d = .02/-0.09/-0.15$) - Refusal efficacy ($d = -0.15/-0.08/0.15$) - Hypothetical alcohol resistance ($d = 0.06/0.16/0.16$) - Number of strategies ($d = 0.13/0.04/0.15$)	- Past month substance use: The increase from baseline to wave 6 (ie. 8th grade) in the prevalence of past month's use reported by control condition students (from 43% to 56%) was significantly less than that reported by students receiving the kiR-AE (keepin' it real-acculturation enhanced) curriculum in only the 5th grade (from 29% to 78%).

Authors (Year)	Dependent Variables	Measures	Outcomes of Measures (Quantitative, if possible, otherwise qualitative provided)	Author Description of Findings
	resistance strategies -Personal anti-drug norms (ie. rejection of peer substance use) -Descriptive substance use norms (ie. perception of prevalence of peer substance use) -Positive substance use expectancies		-Personal anti-drug norm ($d = -0.06/-0.11^*/-0.06$) -Descriptive substance use norms (ie. views that peers are using substances) ($d = -0.02/-0.26/-0.21$) -Positive substance use expectancies ($d = 0.07/0.11/-0.34$)	-Lifetime substance use: The increase from baseline to wave 6 (ie. 8th grade) reported by control students did not differ significantly from that reported by intervention students, except in the case of students receiving the kiR-AE version in only the 5th grade. Students in that condition increased in their lifetime prevalence of substance use (from 23% to 71%) significantly more than did control condition students (from 30% to 63%) Students receiving the curriculum versions in the 7th grade or in both the 5th and 7th grades did not appear different from the control group with regards to recent substance use.
Graves et al. (2017)	-Social-emotional knowledge -Social-emotional assets -Externalizing behaviors	-Strong Start Content Knowledge Assessment (SSCK) -Social-Emotional Assets and Resiliency Scales (SEARS) -BASC-2	-Social-emotional knowledge: 16% increase, d not calculable -Social-emotional assets: self-regulation: $d = 0.99$ large self-competence: $d = 1.38$ large empathy: $d = 0.47$ small responsibility: $d = 0.10$ small -Externalizing behaviors ($d = 0.36$ small)	-Social-emotional knowledge: "From pretest to posttest, the students' scores increased on the Strong Start Content Knowledge assessment. Between pre- and posttest, students demonstrated an increase in their social-emotional knowledge. Specifically the group average was 68% correct on the pretest and 84% on the posttest."

Authors (Year)	Dependent Variables	Measures	Outcomes of Measures (Quantitative, if possible, otherwise qualitative provided)	Author Description of Findings
				-Social-emotional assets: “Students’ self regulation and self-competency improved over the course of the intervention.” Empathy and responsibility did not have statistically significant differences between before and after scores of the intervention. -Externalizing behaviors: “There were not any statistically significant decreases in teacher rated externalizing problems...the change in student’s aggression as measured by effect sizes was not practically meaningful.”
Hecht et al. (2018)	-Youth lifetime alcohol use -Youth lifetime cigarette use -Youth lifetime marijuana use -Youth lifetime chewing tobacco use	-4 items adapted from Hansen and Graham, 1991)	▲Wave 1-Baseline: (d = high quality adapted vs control/low quality adapted vs control) -Alcohol use (d = 0.07/0.145) -Cigarette use (d = 0.08/0.16) -Marijuana use (d = 0.068/0.127) -Chewing tobacco use (d = -0.104/-0.117) ▲Wave 4 - 24 months after implementation): (d =high quality adapted vs control/low quality adapted vs control) -Alcohol use (d = -.14/-0.11) -Cigarette use (d = -.20/-0.13)* -Marijuana use (d = -0.007/0.0569) -Chewing tobacco (d = -.169/-0.20)	-Youth in both standard and rural adaptation programs reported significantly less cigarette use than those in the control condition. -Programs with high quality delivery in the multicultural/rural conditions were in the desired direction but not statistically significant. -Youth in both high- and low-quality delivery of the adapted/rural condition reported significantly less cigarette use than those in the control condition. -All the other effects for high-quality adapted/rural curriculum were in the desired direction but not statistically significant.

Authors (Year)	Dependent Variables	Measures	Outcomes of Measures (Quantitative, if possible, otherwise qualitative provided)	Author Description of Findings
				-Effects for low-quality adapted/rural curriculum were not statistically significant but were in the desired direction for alcohol use and chewing tobacco use -Youth in the high-quality, nonadapted/urban condition reported less marijuana use than those in the low quality, non-adapted/urban condition but neither high nor low quality delivery of the non-adapted curriculum produced substance use outcomes that differed significantly from the control.
Hopson (2008)	-Past month alcohol use -Past month marijuana use -Intentions to accept alcohol, -Intentions to accept marijuana	-Texas School Survey of Substance Use (3 items adapted for past month alcohol use) -Texas School Survey of Substance Use (1 item adapted for past month marijuana use) -Drug Resistance Strategies (DRS) Project (2 items adapted for intentions to accept alcohol)	-Alcohol use by group ($\eta_p^2 = 0.097$) -Alcohol use by group x time ($\eta_p^2 = 0.108$)* -Alcohol use by group x time x age ($\eta_p^2 = 0.295$)** -Alcohol use by time ($\eta_p^2 = 0.164$) ** -Marijuana use by group ($\eta_p^2 = 0.007$) -Marijuana use by group x time ($\eta_p^2 = 0.042$) -Marijuana use by group x time x age ($\eta_p^2 = 0.057$) -Marijuana use by time (partial eta squared = 0.044) -Intentions to accept alcohol by group ($\eta_p^2 = 0.24$) -Intentions to accept alcohol by group x time ($\eta_p^2 = 0.23$) ***	-There was a significant difference in alcohol use between groups over time ($F(2, 60) = 3.637, p < .05$). -Alcohol use accounted for 10.8% of the difference in group means over time. -Alcohol use: The age difference between groups may have influenced the difference in mean scores over time. -Alcohol use decreased over time for the experimental group and remained fairly stable for the comparison group, indicating that the experimental group reported significantly greater decreases in

Authors (Year)	Dependent Variables	Measures	Outcomes of Measures (Quantitative, if possible, otherwise qualitative provided)	Author Description of Findings
		-Drug Resistance Strategies (DRS) Project (1 item adapted for intentions to accept marijuana)	-Intentions to accept by alcohol group x time x age ($\eta_p^2 = 0.142$) -Intentions to accept alcohol by time ($\eta_p^2 = 0.037$) -Intentions to accept marijuana by group ($\eta_p^2 = 0.001$) -Intentions to accept marijuana by group x time ($\eta_p^2 = 0.034$) -Intentions to accept marijuana by group x time x age ($\eta_p^2 = 0.059$) -Intentions to accept marijuana by time ($\eta_p^2 = 0.004$)	alcohol use than the comparison group -The groups did not differ in their marijuana use over time -Although the difference was not significant, the experimental group's marijuana use decreased steadily over time, while the comparison group's use decreased slightly and then increased -Intentions to accept alcohol: Experimental group was more likely to disagree that they would accept an offer for alcohol than the comparison group over time -Intentions to accept marijuana: The experimental group scores increased over time while the scores for the comparison group decreased. Since the analysis suffers from lack of power due to small sample and violation of assumptions, the trend warrants further study
Kurtz (2020)	-Students' internalizing behaviors	- The Behavior Intervention Monitoring Assessment System 2 (BIMAS-2)	Negative affect scores: from pre intervention to post intervention -Class wide condition: $d = -1.23$ -Small group condition: $d = -2.87$ -Receiving special education: $d = -2.87$ -General education: $d = -1.14$	-“Decreases in Negative Affect T scores suggest improvement in symptoms/functioning while increases in scores may reflect worsening symptoms (McDougal et al., 2011). Negative Affect T scores

Authors (Year)	Dependent Variables	Measures	Outcomes of Measures (Quantitative, if possible, otherwise qualitative provided)	Author Description of Findings
				decreased from pre-intervention to post-intervention for both the classwide, pre-intervention: $M = 65.42$, $SD = 5.71$ post-intervention: $M = 58.42$, $SD = 7.35$, and small group conditions, pre-intervention: $M = 66.68$, $SD = 5.49$, post-intervention: $M = 56.63$, $SD = 9.49$, suggesting teachers rated their students as having fewer problem behaviors related to Negative Affect after the intervention."
Martinez (2011)	-Educational aspirations -Adjustment (psychological) -Anger management -Self-esteem -Ethnic pride -Self-reported peer attachments -Gang involvement	- Junior High life Experiences Survey (JHLES) -Survey of Children's Exposure to Community Violence (SECV-SR) -Educational Aspirations (researcher designed) -Trauma Symptom Checklist for children (TSCC) -Anger management (researcher designed + vignettes from	-Educational aspirations: $[F(1,29) = 1.42, p = .24]$ -Psychological adjustment: $[F(1,29) = .01, p = .94]$ -Anger management: "From pre-group to post-group, participants said they would be less likely to walk away and more likely to try to find out why people were fighting." -Self-esteem: $[F(1,29) = 2.11, p = .16]$ -Ethnic pride: $[F(1,29) = 2.11, p = .16]$ -Peer attachment: Elementary students: $[F(1, 16) = 6.06, p = .026]$; Junior high school students: $[F(1, 13) = .49, p = .50]$ " -Gang involvement: $[F(1,29) = 1.55, p = .22]$	-Educational aspirations: no significant pre-group to post-group change -Psychological adjustment: no significant pre-post group change -Anger management: provided in Outcomes of Measures column -Self-esteem: "No significant changes from pre- to post-group in self-esteem...Participants tended to report they agree to feeling self-esteem." -Ethnic pride: There were no significant changes from pre- to post-group in ethnic pride. Participants tended to report that they agree to feeling ethnic pride

Authors (Year)	Dependent Variables	Measures	Outcomes of Measures (Quantitative, if possible, otherwise qualitative provided)	Author Description of Findings
		Slaby & Wildon- Brewer 1992) -Rosenberge Self- Esteem Scale -Multigroup Ethnic Identity Measure (MEIM) -Peer Scale of the Inventory of Parent and Peer Attachment (IPPA) -Gang Involvement (researcher designed)		-Peer attachment: "Participants in elementary school showed an improvement in their reported peer attachment from pre to post-group...However, there was no significant change for junior high school participants" -Gang involvement: "Participants tended to report that they did not find it that necessary to belong to a gang. There were no significant changes from pre- to post-group in gang involvement"
Moran et al. (2019)	-Knowledge and confidence to intervene -Using "stealing the show" strategy -Using "turning it over" strategy -Using "accompanying others" strategy -Using "coaching compassion" strategy	-Student-Advocates pre- and post-scale -Use of STAC strategies (each strategy measured by 1, 5 point-Likert item)	Results from full sample: -Knowledge and confidence to intervene by time ($\eta_p^2 = .68$) *** -Knowledge and confidence to intervene by time x ethnicity ($\eta_p^2 = .02$) -Using "stealing the show" strategy (white vs Hispanic student use: $\phi = .39$)* -Using "turning it over" strategy (white vs Hispanic student use: $\phi = .02$) -Using "accompanying others" strategy (white vs Hispanic student use: $\phi = .02$) -Using "coaching compassion" strategy (white vs Hispanic student use: $\phi = .02$)	- Results indicated Hispanic students used the strategy "stealing the show" more frequently than White students, with no significant differences between White and Hispanic students for the other three strategies

Note. Where possible, author language was used. Effect sizes were extracted from papers. ▲ symbolizes calculations made by the authors of this systematic review. * indicates $p < .05$; ** indicates $p < .01$; *** indicates $p < .001$ ϕ indicates partial eta squared, which "" means the

proportion of the variance that is attributable to the effect of the dependent variable" (Hopson, 2008, p. 136) d = indicates the Cohen's D value, which demonstrates the number of standard deviations between the two means. In Cramer and Castro-Olivio (2015), the effect size type calculated by the authors was not reported, thus the authors' language of "ES" was used in this review.

Table S10*Social Validity and Outcomes*

Authors (Year)	Social Validity Construct(s)	Measure	Outcome
Castro-Olivo and Merrell (2012)	-Social validity (student): acceptability and satisfaction -Social validity (teacher): perceived acceptability, satisfaction, feasibility	-Social validity: student report (researcher developed, 7-item scale) -Social validity: teacher report (researcher developed, 16-item scale)	- Student acceptability: Not Reported - Student satisfaction: highly satisfied (80%+) - Teacher perceived acceptability: both said students liked the intervention - Teacher satisfaction: very satisfied - Teacher feasibility: "good" to "very good"
Castro-Olivo (2014)	-Acceptability -Satisfaction	Social validity: student report (researcher developed, 9-item scale)	- Acceptability: (high, 84.4%+ endorsed statements) - Satisfaction: Not Reported
Castro-Olivo et al. (2022)	Not reported	BERS-2 (restructured) (student completed)	Not reported
Cramer and Castro-Olivo (2015)	-Acceptability -Satisfaction -Overall social validity	9-item measure based on Wolf's (1978) definition of social validity. Scale was validated in previous studies examining a culturally adapted version of Strong Teens (student only)	- Acceptability: Not reported as expected - Satisfaction: Not reported as expected - Overall social validity: Not reported as expected -In general, high ratings on all 9 items. Ranged from 80-100% agreement (SD = 0.07).
Elek et al. (2010)	MISSING	MISSING	MISSING
Graves et al. (2017)	- Social significance, social importance, social appropriateness -Teacher view of social validity	-Student social validity questionnaire -Teacher survey -Teacher interview	- Social significance, social importance, social appropriateness: high endorsements of all items - Teacher view of social validity: viewed the intervention as a good first step, but needed

Authors (Year)	Social Validity Construct(s)	Measure	Outcome
			a stronger focus on Black boys, resilience, and needed more time
Hecht et al. (2018)	MISSING	MISSING	MISSING
Hopson (2008)	MISSING	MISSING	MISSING
Kurtz (2020)	-Teachers' acceptability and likelihood of continued implementation of the adapted intervention	-The Usage Rating Profile Intervention-Revised (URP-IR), teacher rated	Max score = 6 -Acceptability ($M = 5.32, SD = 0.96$) -Understanding ($M = 3.92, SD = 0.95$) -Home school collaboration ($M = 4.86, SD = 0.91$) -Feasibility ($M = 4.78, SD = 0.79$) -System climate ($M = 5.50, SD = 0.56$) -System support ($M = 4.19, SD = 1.25$)
Martinez (2011)	-Overall perceived effect -Overall program evaluation	-Perceived effect scale (student rated) Likert Scale: 1 = Not at all; 5 = Very helpful -Post interview (students only)	-Overall perceived effect: $N = 31, M = 3.56 (.84)$ -Overall program evaluation ($N = 31, M = 4.54 (.69)$) -Group interview: many of the participants felt an overall appreciation for the program and how it helped them; help with anger management was listed as the area in which they felt most helped. -Participants reported that they learned most about not bullying others, anger management, educational aspirations, expressing their feelings and one student mentioned culture and racism.

Authors (Year)	Social Validity Construct(s)	Measure	Outcome
Moran et al. (2019)	Students' experiences participating in the culturally adapted STAC program and using the STAC strategies to intervene in bully situation in a predominantly white and Hispanic middle school located in a low-income rural community	- Participants' experience (student interview)	Results only from qualitative sample: - 47.8% of students spoke about an increase in awareness about bullying and reported using all four strategies to intervene across multiple contexts -39% of students intervened using the STAC strategies despite their fears -56.5% of students reported feeling good about themselves, confident, and feeling like they could better recognize bullying to make a difference at their school after using STAC strategies -43.4% of students reported that since participating in the training and using STAC strategies they are making new friends, building stronger relationships, and becoming more discerning in friendships

Note. "Acceptability" and "satisfaction" may have been the general ways authors were defining social validity, however, for the most clearway of coding the two, the two were separated. MISSING: The information is not mentioned in the article at all; NOT EXAMINED: The authors explicitly mention they did not analyze a variable; NOT REPORTED: The authors had a measure, but did not report findings; NOT REPORTED AS EXPECTED: The way the data were reported were not in the same unit of study as other variables (e.g. reports demographic for the entire school instead of for the group)