

ORIGINAL ARTICLE

Scaling up effective juvenile delinquency programs by focusing on change levers: Evidence from a large meta-analysis

David B. Wilson¹  | Mark W. Lipsey²

¹Department of Criminology, Law and Society, George Mason University, Fairfax, Virginia, USA

²Department of Human and Organizational Development, Vanderbilt University, Nashville, Tennessee, USA

Correspondence

David B. Wilson, Department of Criminology, Law and Society, George Mason University 4400 University Drive, MS 4F4, Fairfax, VA 22030, USA.
Email: dwilsonb@gmu.edu

A prior version of this work was presented as a keynote address at the *Applied Research in Crime and Justice Conference*, Sydney, Australia, on February 13, 2019.

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Funding information

Creation and updating of the database used in this study were supported by grants from the Office of Juvenile Justice and Delinquency Prevention 2001-JN-FX-0008, the National Institute of Mental Health MH57766, MH42694,

Abstract

Research summary: The primary outcome desired for juvenile delinquency programs is the cessation of delinquent and related problematic behaviors. However, this outcome is almost always pursued by attempting to change intermediate outcomes, such as family functioning, improved mental health, or peer relations. We can conceptualize intermediate outcomes that are related to reduced delinquency as change levers for effective intervention. A large meta-analysis identified several school-related change levers, including school engagement (i.e., improved attendance and reduced truancy), nondelinquent problem behaviors, and attitudes about school and teachers. In addition, family functioning and reducing substance use were also effective change levers. In contrast, effects on youth getting/keeping a job, peer relationships, and academic achievement were not associated with reduced delinquency.

Policy implications: Only a small percentage of rehabilitative programs provided to youth involved in the juvenile justice system have been established as evidence based. Moreover, there are constraints on what

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *Criminology & Public Policy* published by Wiley Periodicals LLC on behalf of American Society of Criminology.

and MH39958, the W. T. Grant Foundation, and the Russell Sage Foundation.

local policy makers and practitioners can do regarding the selection, adoption, and implementation of programs from the available lists of evidence-based programs. Adopting programs that focus on effective change levers and avoiding those that concentrate on ineffective ones has the potential to increase the likelihood that a local agency is engaged in effective programming. Based on our data, programs known to improve family functioning, attachment to and involvement in schooling, and reducing substance use are justified by the change lever evidence, even if these programs' effectiveness in reducing delinquency has not been directly proven. In contrast, programs focusing on vocational skills, academic achievement, and peer relations are less likely to be beneficial. Furthermore, a change lever perspective can help frontline staff select appropriate programs for different juvenile offenders and focus their quality control efforts on those aspects of a program that are likely to be essential to maintaining effectiveness.

KEYWORDS

change levers, delinquency, mediators, meta-analysis, program effectiveness

Evidence-based programming for juvenile delinquency focuses on the effectiveness of specific programs and program types. For example, the CrimeSolutions.gov (crimesolutions.ojp.gov) registry lists programs and practices of this sort. Programs are defined as "a specific set of activities carried out according to guidelines to achieve a defined purpose," and practices are defined as "a general category of programs, strategies, or procedures that share similar characteristics with regard to the issues they address and how they address them." This paper takes a complementary approach and focuses on the potential change mechanisms or intermediate outcomes through which programs attempt to achieve the goal of reduced delinquency.

The only direct control of delinquency is incapacitation, which has significant costs, can make matters worse, and is not appropriate for most youthful offenders. Most delinquency programs, therefore, must work indirectly by changing something that, in turn, affects delinquency, for example, by increasing or improving something desirable that is presumed to be negatively related to delinquency, such as school attendance, interpersonal skills, prosocial attitudes, or family functioning. One advantage to focusing on these intermediate outcomes is that there are relatively few theoretically plausible change pathways that affect delinquent behavior.

In contrast, the number of distinct intervention programs that could be created to address delinquency is limitless, and a considerable number have already been developed. For example, as of April 2023 CrimeSolutions.gov lists 689 programs focused on juveniles. That list represents a lot of programs to evaluate and consider for adoption. And determining the effectiveness of each

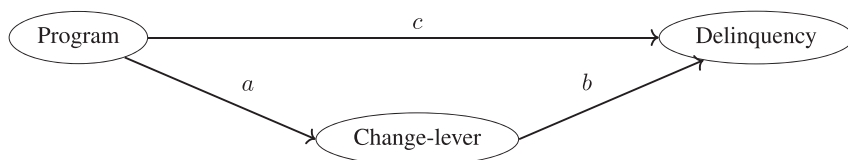


FIGURE 1 Visualization of a change lever.

of these programs is a massive social science undertaking. The vast majority of these programs have had too few evaluations to establish with any degree of confidence whether they are effective in routine applications. This plethora of programs is an inefficient use of research resources. Better evidence about potential mediating change mechanisms could redirect attention to how programs achieve their effects and allow whole families of programs to be categorized according to the primary mediating outcomes addressed and their impact on those outcomes.

We call these potential mediating mechanisms *change levers* to reflect their role as the intermediate outcomes a program acts on to leverage change toward the ultimate goal of reduced delinquency. Actual mediation assumes that change in a mediator *causes* a change in a target outcome, as illustrated in Figure 1. For the change lever to be an actual causal mechanism, paths *a* and *b* must account for a meaningful portion of path *c*. However, to demonstrate that a candidate mediator is a causal mechanism, it must be shown that manipulating it individually and directly produces a change in the outcome (Kazdin, 2007). While delinquency interventions often target a potential mediator (e.g., social skills, self-esteem), few are so focused that effects are unlikely on other potential mediators or those correlated with the one targeted, and it is not unusual for researchers to measure effects on nondelinquency outcomes other than those most directly targeted. We thus view change levers as variables with mediational characteristics, that is, outcomes empirically related to delinquency reduction, but do not assume that they are necessarily true causal mediators or the only such variables affected by an effective intervention. Those empirical relationships, nonetheless, justify a programming focus on identifiable change levers as especially plausible routes to effective intervention.

Notice that these change levers are not simply risk factors predictive of delinquency. Our focus is on intervention-induced effects on a change lever. These may or may not correspond to risk factors for delinquency. Just because a variable is predictive of, or even causally related to, youth becoming delinquent does not necessarily mean that changing that variable after a juvenile has engaged in delinquency will transition them away from crime. Stated more simply, what helps youth transition out of delinquency may differ from what leads them into it. This point makes the change levers framework different from the RNR (risk-needs-responsivity) approach with its focus on risk factors and criminogenic needs.

Within the RNR model, criminogenic needs are the target change levers that treatment programs should address. The main criminogenic factors listed by Andrews et al. (2006) were established through an examination of predictors of recidivism (Gendreau et al., 1996: they call these the “BIG 8”): (1) history of antisocial behavior, (2) antisocial personality, (3) antisocial values and attitudes, (4) antisocial peers, (5) substance abuse, (6) dysfunctional family relations, (7) education/employment, and (8) leisure. Several of these factors are static, that is, unchangeable features of an individual and, as such, unsuitable as a change lever. Recognizing this problem, Andrew and colleagues focus rehabilitation efforts on dynamic needs that flow from these criminogenic factors, for example, problem-solving skills, self-management skills, anger management, coping skills, antisocial cognitions, adopting an anti-criminal identity, reducing association with criminal others, enhancing work performance, reducing substance abuse, and increasing

alternatives to drug abuse (Andrews et al., 2006). Each of these is a potential change lever within our framework. However, a strong correlation between a risk factor or dynamic need and criminal offending does not establish that changing that risk factor will also change offending. Meta-analyses have established that rehabilitation programs that use the RNR model are generally more effective than those that do not (Andrews et al., 2006; Koehler et al., 2013). Missing from much of this work, however, is direct evidence establishing the empirical link between a reduction in dynamic needs and recidivism.

The connection of risk and protective factors to program mediators is not unique to the RNR model. For example, in a paper discussing various issues related to evidence-based juvenile justice programs, Elliott et al. (2020) argued for the value of systematic reviews identifying “the chain of logic that links the intervention to the outcome of interest” (p. 1321). That paper conceptualizes these links as the “change strategies employed to modify risk and protective factors” (p. 1321). Our conceptual shift from the risk and protective factors of developmental criminology to change levers more generally may seem minor but broadens the range of potential change levers to mediating variables that may not be causally related to the initiation of delinquency, but may be causally associated with desistance from it.

One approach to identifying change levers is to examine the relationship between effects on intermediate outcomes and those on recidivism in the existing base of intervention studies with juvenile offenders. The magnitude of the relationships between those intermediate effects and recidivism effects provides some indication of which intermediate variables have the most promise as effective change levers and are most worthy of further attention in research and practice. The analyses we present use data from a large meta-analysis of the effects of juvenile delinquency programs to examine the relationship between the programs’ effectiveness in changing intermediate outcomes and their ability to reduce the recidivism of the participating juveniles. Prior work with a much earlier version of this meta-analysis showed that such analysis could be informative (Lipsey, 1992a, 1995).

1 | METHODS

We used data from the second author’s large meta-analysis of research on the effects of juvenile delinquency programs to identify potentially effective change levers. This database includes 548 independent study samples drawn from research reports released between 1958 and 2015, further described in Lipsey (2009). Lipsey and his team coded treatment effects on delinquency plus effects on all the other outcomes reported in each study for which an effect size could be computed, for example, mental health, interpersonal adjustment, vocational, and school-related measures. This extensive coding of effects on nondelinquency outcomes makes possible the analyses presented below that examined whether treatment effects on those intermediate outcomes were related to treatment effects on delinquency.

1.1 | Systematic search

The systematic search for studies eligible for this meta-analysis has been extensive. It began in the mid-1980s and has been updated numerous times. The initial search procedure is most fully described in Lipsey (1992a) and was replicated in the later updates. It included an examination of prior literature reviews and meta-analyses, a keyword search of multiple relevant bibliographic

databases, manual searches of journals that often publish eligible studies, and citations within the reports identified from these sources that were retrieved and screened for eligibility. Candidate study reports were retrieved from the university library, interlibrary loan, purchase of microfiche and books, and requests to authors.

1.2 | Criteria for study eligibility and selection for analysis

Studies must have met the following eligibility criteria to be included in the original meta-analysis: (a) the research was conducted in an English-speaking country and reported in English, (b) the interventions were designed to reduce delinquency, (c) the juveniles studied were between 12 and 21 years of age and selected for delinquency or similar antisocial behavior, (d) at least one delinquency outcome variable was measured, (e) the outcomes of the intervention program were directly compared with those of a control group of similar juveniles who did not receive the intervention. For additional details on these eligibility criteria, see Lipsey (1992b).

Of the 548 studies in this meta-analysis, 197 reported nondelinquency effect sizes. We restricted our analyses to effects on nondelinquency outcomes reported in at least 20 studies to ensure a sufficient number for meaningful analysis. This selection resulted in 153 studies that included 173 unique study samples for use in the analysis when independent samples for different jurisdictions or treatment groups were counted. Note that multiple publications based on the same study data were coded as a single study, avoiding duplications in our data. The references for these studies are included in the [supplemental file](#) for this article.

1.3 | Categorization of nondelinquency constructs

The study coding for the original meta-analysis categorized the constructs represented in the nondelinquency outcome measures into a hierarchical scheme comprised of a small number of broad categories under which subcategories were nested. The development of this hierarchy emerged through an interactive process of coding studies in the early phase of this meta-analysis and developing categories for newly identified nondelinquency outcomes as they were encountered. Refinements to this categorization hierarchy have been made based a consideration of cohesiveness across studies of the measures in a common category. As with all coding for this meta-analysis, reliability was ensured through double coding by two coders and a reconciliation process to resolve any coding differences. Furthermore, to ensure both the accuracy of this categorization and its suitability for the change-lever framework, we reexamined the description of each nondelinquency measure for which an effect size was available and adjusted the categorization in cases where that was judged to produce a better fit for the purposes of the present study. Only a few nondelinquency outcomes were recategorized as result of this process. The resulting categorization scheme with examples of the measures represented in each category is as follows.

1.3.1 | Psychological

- Attitudes about delinquency/prosocial conduct; for example, respect for property of others, attitudes toward antisocial behavior, antisocial thinking, criminal beliefs and attitudes,

socially acceptable goals, rationalizations for criminal behavior, susceptibility to deviant peer influence.

- Personality traits; for example, rigidity and closed-mindedness, empathy, locus of control, responsibility, defensiveness, impulsivity, callous and unemotional, self-control, inhibition, state-trait anger.
- Self-esteem/self-concept; for example, assessment of self-worth, favorable view of self, positive self-concept, self-esteem, self-image, self-confidence, self-satisfaction, self-criticism, self-efficacy.
- Internalizing/anxiety/depression; for example, internalizing, suppression of aggression, alienation, depressive symptoms, anxiety symptoms, emotional regulation, emotional health.
- Substance use; for example, drug use, illicit drug use, chemical abuse, substance abuse, alcohol, binge drinking, marijuana, hard drugs, cocaine, positive drug test, gateway drug use, polydrug use, abstinence.

1.3.2 | Interpersonal

- Family functioning; for example, family relations, family environment, family dissension, relationship with parents, parental rejection, family interaction, family cohesion, parental supervision, intrafamily violence, abuse, family communications, family conflict, family climate.
- Peer relations; for example, peer relations composite, delinquent associates, friendliness rating, social anxiety rating, relations with friends, peer pressure, negative peer influence, delinquent friends, peer support.
- Social skills; for example, social skills rating, social competence with peers, social adjustment, interpersonal competency, social coping skills, social problems, social maladjustment, social perspective taking.

1.3.3 | School

- Attendance; for example, school absence, school tardiness, number days absent, attending school on at least 80% of school days, unexcused absences.
- Nondelinquent behavior resulting in discipline; for example, school suspensions, number of school disciplinary incidents, behavioral problems, discipline referrals, days suspended, school conduct.
- Attitudes about school/teachers; for example, attachment to school, classroom engagement, importance of school, school commitment.
- Dropping out/graduating; for example, school dropout, remained in school, dropout rate, high school diploma, persist all year, % graduated, proportion classes passed, % withdrew prematurely, promoted, highest grade completed, received GED or high school diploma, still in school.
- Academic achievement; for example, reading grade level, reading comprehension, academic performance, academic success, grades, GPA, school credits, proportion of classes passed, failing grade average, grade retention.

1.3.4 | Vocational

- Employment status; for example, gets/keeps job, job seeking, employment placement, full-time employment, same job for six months, rate of employment, % unemployed, weeks employed, currently working, months worked, lost job/fired.

Note that these categories vary with regard to the conceptual heterogeneity of the outcome measures included. Some are quite coherent (e.g., social skills, school attendance) while others are more diverse (e.g., personality traits). That variation is likely to have some influence on the correlations with delinquency outcomes.

1.4 | Effect sizes

Hedges' *g* standardized mean difference effect size was used for all outcome constructs. For dichotomous outcomes, a logged odds ratio was calculated and converted to Hedges' *g* using the logit method described in Sanchez-Meca et al. (2003) and Hasselblad and Hedges (1995). For continuous outcomes, means and standard deviations were used when available. In all other cases, the estimation method was based on formulas presented by Lipsey and Wilson (2001). More than three-fourths of the studies used either random assignment, matched designs, or waitlist controls. In addition, we adjusted effect sizes for baseline differences when possible. This adjustment was rarely possible for delinquency outcomes (occurring for only 21 of 173 effect sizes) given that this can only happen when there is the same measure of delinquent behavior reported at both pretest and posttest. In contrast, we could compute a baseline adjusted effect size for 17% of the nondelinquency outcomes (74 of 435). For effect sizes based on means, these baseline-adjusted effect sizes were computed as a difference-in-differences of the means and standardized using the posttest standard deviations. For effect sizes based on logged odds ratios, the baseline logged odds ratio was subtracted from the posttest effect size before conversion to Hedges' *g*. Both delinquency and nondelinquency effect sizes were coded such that positive values reflect positive changes (i.e., less delinquency, improved family functioning, etc.) and negative values reflect negative changes (i.e., more delinquency, less desirable nondelinquency outcomes).

Studies often reported multiple delinquency effect sizes. For the analyses reported here, we preferred effect sizes (a) based on the largest sample size (i.e., effect sizes with the minimum amount of attrition), (b) representing general offending (not a specific offense type), (c) measured either at 52 weeks posttest or the effect size closest to this time point, and (d) based on arrest or police contact, court contact, institutionalization, and unofficial measures of delinquency, such as self- or other-report (in that order). For most studies (77%), this selection resulted in a single effect size per independent study sample. For the remaining studies with multiple preferred effect sizes per independent study sample, those effect sizes were aggregated using the method detailed by Pustejovsky and Chen (2024). This approach produces results that are equivalent to the robust variance estimation method of Hedges et al. (2010) for the types of models we estimated but does so by first producing a single effect size per unique study sample with an associated standard error that assumes dependence among the aggregated effect sizes.

The selection process for nondelinquency effect sizes was more straightforward. First, we selected those effect sizes measured at posttest, that is, after treatment completion. Most nondelinquency effect sizes were measured only at posttest (89%), and all had a posttest effect

size. When there were multiple preferred effect sizes within a construct category for any independent sample, we aggregated them, again using Pustejovsky's method, thus producing a single effect size and associated standard error per construct of interest per unique study sample.

Effect sizes greater than an absolute value of 1.25 were Winsorized to 1.25 (or -1.25 if negative). This was done to handle a small number of extreme effect sizes and reduce their influence on the slope of the regression lines estimated in the analyses below.

1.5 | Analyses

The analyses were performed in **R** using the *metafor* package (Viechtbauer, 2010). We assumed a random effects model and estimated the random effect variance component, τ^2 , via restricted maximum likelihood. We regressed delinquency effect sizes onto nondelinquency effect sizes using the standard error of the delinquency effect size as the basis for the inverse variance weight. These models produced a regression coefficient, labeled B_1 in the tables, that reflects the relationship between the nondelinquency and delinquency effect sizes. We also estimated an unweighted version of B_1 as a sensitivity analysis on the influence of the inverse variance weights. The intercept is labeled B_0 in these tables.

As a further sensitivity analysis, we examined the relationship between the nondelinquency constructs and delinquency while including covariates that accounted for differences across studies on selected methodological features. Because the relevant methodological features were different for the nondelinquency and delinquency measures, this analysis was conducted in two stages. First, the delinquency effect sizes were predicted from a set of variables that included (a) whether a study was randomized, (b) whether the delinquency measure was based on official records, and (c) the log of the number of weeks posttest captured by the delinquency measure. The residuals from this regression (i.e., the part of the delinquency effect sizes that could not be explained by the covariates) were centered around the mean for randomized studies with an official measure of delinquency measured at 52 weeks posttest to yield method-adjusted delinquency effect sizes. Similarly, the nondelinquency effect sizes were adjusted for (a) whether the study was randomized, (b) whether the nondelinquency outcome was a standardized measure (e.g., a published instrument with known psychometric properties), and (c) whether the measure was based on the youth's self-report. The residuals from this analysis were also collected and centered around the mean for randomized studies yielding method-adjusted nondelinquency effect sizes. In the second step, a single random-effects meta-regression model was fit for the method-adjusted delinquency effect sizes regressed on the method-adjusted nondelinquency effect sizes plus a set of dummy codes that identified the respective nondelinquency constructs.

2 | RESULTS

Table 1 presents descriptive statistics for the 173 independent study samples from the 153 studies included in these analyses. From this table, we can see that over half of the manuscripts were from the grey literature. More specifically, roughly a third of the manuscripts were technical reports (33.5%), a fifth were dissertations or master's theses (20.2%), less than 5% were books (4.6%), and less than 1% were conference papers. This high percentage of grey-literature manuscripts reduces the likelihood of our results being overly influenced by publication-selection bias. However, it

TABLE 1 Selected characteristics of included studies.

Variable	Category	Frequency	Percent
Publication type	Journal-Book chapter	71	41.0
	Technical report	58	33.5
	Thesis-dissertation	35	20.2
	Book	8	4.6
	Conference paper	1	0.6
Decade Published	1960s	14	8.2
	1970s	53	31.0
	1980s	52	30.4
	1990s	15	8.8
	2000s	25	14.6
	2010s	12	7.0
Country of study	United States	158	91.3
	United Kingdom	8	4.6
	Canada	6	3.5
	Other	1	0.6
Research design	Random assignment	104	60.8
	Nonrandom, convenience	35	20.5
	Nonrandom, matching	24	14.0
	Other	4	2.3
	Wait-list	4	2.3
Type of delinquency measure	Arrest/police contact	73	42.2
	Other	49	28.3
	Court contact	34	19.7
	Institutionalization	11	6.4
	Antisocial behavior	6	3.5
Source of delinquency measure	Official records	127	73.4
	Self-report	33	19.1
	Other-report	7	4.0
	Other	6	3.5
Type of nondelinquency measure	Survey type items	142	32.6
	Standardized	116	26.7
	Archival report	93	21.4
	Other	84	19.3
Source of nondelinquency measure	Juveniles	246	56.6
	Records/archives	93	21.4
	Other	72	16.6
	Nonresearcher observers	21	4.8
	Researchers	3	0.7

(Continues)

TABLE 1 (Continued)

Variable	Category	Frequency	Percent
Months captured by del. measure posttreatment	<6-months	53	30.6
	6-months to <12-months	52	30.1
	12-months	40	23.1
	>12-months to 24-months	13	7.5
	>24-months to 36-months	5	2.9
	>36-months	2	1.2
	Missing	8	4.6
Sample size	<50	55	31.8
	50–99	33	19.1
	100–199	40	23.1
	200–499	29	16.8
	500–999	10	5.8
	1000+	6	3.5
Sex distribution	Some males (<50%)	6	3.5
	Mostly males (= or >50%)	10	5.8
	All males (>95%)	74	42.8
	Cannot tell	83	48.0s

Note: There are 173 unique study samples from 153 studies.

remains plausible that nondelinquency outcomes were more likely to be reported, on average, when they were statistically significant.

Most of this research was conducted before 2000, representing slightly over three-fourths of the studies (78.4%). Slightly less than 10% were published in 2010 or later. Most of these studies were conducted in the United States (90.7%), with a small percentage coming from the United Kingdom and Canada.

Over half of these studies used a random assignment to conditions design (60.8%). Most of the remaining studies used nonrandom designs with either a basic nonequivalent comparison group design without matching (20.5%) or some form of matching (14.0%). There were a small number of wait-list control designs or some other quasi-experimental design type.

Ideally, we want a large range in the ability of these interventions to produce a change in the nondelinquency outcomes. Otherwise, it will be difficult to detect a meaningful relationship between change in these constructs and change in delinquency. Table 2 presents the mean effect size and associated statistics for outcome constructs reported in 20 or more studies. These mean effect sizes range from essentially a null value for family functioning and substance use to a moderately positive effect on nondelinquent school behavior that resulted in discipline ($d = 0.35$). In addition, the mean effect size for nine of these 14 constructs is statistically significant at a conventional level (i.e., the lower 95% confidence interval is greater than zero).

The interquartile range (IQR) is of greater interest to our purpose. It ranges from a low of 0.37 to a high of 0.75. The small range in effect sizes for some of these constructs, such as substance use and peer relations, may attenuate the observed relationship between these constructs and delinquency, as it is well established that a restricted range in a predictor variable will attenuate the strength of the observed correlation (e.g., Mendoza & Mumford, 1987).

TABLE 2 Mean effect size and related statistics for each nondelinquency effect size category and the delinquency outcome.

Construct	k	Mean g	95% CI		Delinq.			p	τ^2	I^2
			Lower	Upper	IQR	IQR	Q			
Psychological										
Attitudes about delinquency/prosocial conduct	29	0.14	-0.03	0.31	0.56	0.32	118.93	0.0000	0.156	82
Personality traits	35	0.22	0.09	0.35	0.43	0.48	89.55	0.0000	0.088	63
Self-esteem/ self-concept	33	0.21	0.05	0.36	0.54	0.42	168.15	0.0000	0.152	85
Internalizing/anxiety/depression	29	0.13	-0.01	0.27	0.40	0.31	77.75	0.0000	0.083	72
Substance use	36	0.04	-0.06	0.14	0.37	0.26	125.81	0.0000	0.056	81
Interpersonal										
Family functioning	31	0.02	-0.08	0.12	0.39	0.58	62.13	0.0005	0.039	55
Peer relations	33	0.19	0.09	0.28	0.37	0.48	74.18	0.0000	0.040	60
Social skills	21	0.26	0.08	0.44	0.73	0.36	57.36	0.0000	0.103	67
School										
Attendance	50	0.14	0.02	0.27	0.59	0.46	155.18	0.0000	0.118	78
Nondelinquent behavior resulting in discipline	22	0.35	0.16	0.54	0.75	0.43	87.04	0.0000	0.135	83
Attitudes about school/teachers	25	0.20	0.07	0.33	0.61	0.33	79.99	0.0000	0.074	81
Dropping out/graduating	28	0.16	0.01	0.30	0.64	0.33	64.79	0.0001	0.068	74
Academic achievement	32	0.16	0.03	0.29	0.55	0.37	66.98	0.0002	0.062	63
Vocational										
Employment status	31	0.13	-0.01	0.26	0.59	0.50	76.18	0.0000	0.075	78
Delinquency	173	0.19	0.11	0.27	0.44	0.44	709.28	0.0000	0.181	85

Note: All models were estimated under the assumption of random effects using the restricted maximum likelihood (REML) estimator for the random effects variance component (τ^2). k is the number of effect size pairs based on independent study samples included in each analysis. The “Del. IQR” is the IQR for the delinquency effect sizes associated with studies with this nondelinquency effect size.

The final row of the table shows the mean effect size for delinquency. While statistically significant, this effect is small by most standards. However, within the context of criminal justice interventions, it has been argued that such small statistical effects are practically meaningful (Lipsey et al., 2012). Furthermore, the benchmarks established by Cohen are empirically too high in this context, with effect sizes above 0.30 representing the top third and hence “large” by distributional standards (Gies et al., 2023). Finally, the IQR range for delinquency is 0.44 overall, indicating a meaningful range in the effectiveness of these interventions across studies. However, within some constructs, the IQR for the delinquency effect sizes available for analysis is more restricted, such as a 0.26 for substance use.

These distributions are highly heterogeneous with large and statistically significant Q values and I^2 above 55% and as high as 85% for some constructs. This heterogeneity is desirable for the analyses below and indicates that these interventions vary meaningfully in their ability to produce change on nondelinquency constructs. If this were not the case, then it would not make sense to examine the correlation between the nondelinquency effects and the delinquency effects.

2.1 | Relationships between nondelinquency and delinquency effect sizes

Table 3 presents the unstandardized regression coefficients between the nondelinquency and delinquency effect sizes (column B_1). The former is the predictor or independent variable, and the latter is the dependent variable. Because the independent variable is also an effect size and thus has error not incorporated into the model, we do not think that much emphasis should be placed on the standard errors, confidence intervals, I^2 , or τ^2 .

As unstandardized regressions, these coefficients reflect how much delinquency reduction is associated with a given intervention effect on the respective nondelinquency construct. The larger these coefficients, the more likely the nondelinquency construct is an effective change lever for reducing delinquency (note, however, that we cannot draw a direct causal inference from these correlational models). The strongest relationships were for the school-related constructs, with attendance showing the largest coefficient ($B_1 = 0.52$), followed by nondelinquent behavior that resulted in discipline ($B_1 = 0.38$), and attitudes about school/teachers ($B_1 = 0.36$). Within this category, remaining in school (not dropping out or graduating) had a moderate relationship with delinquency outcomes ($B_1 = 0.28$) but fell short of statistical significance. Academic achievement, such as improved grades, had a relatively small relationship with delinquency effects ($B_1 = 0.13$), particularly compared with the other school-related outcomes.

Thus, improving school attendance, preventing dropping out of school, increasing positive attitudes towards school, and reducing in-school problem behaviors are all potential change levers for reducing delinquency among youth. This finding suggests that an important domain of intervention effects from the perspective of reducing delinquency is increased attachment to and participation in school and improved behavior within the school. Focusing on academic performance appears to be less likely to be an effective change lever.

The relationship between school attendance and delinquency is illustrated via the scatter plot in Figure 2. The x -axis of this plot is the effect size for attendance, the y -axis is the effect size for delinquency, and the size of the dots reflects the weight given to each study (larger samples, smaller standard errors). It is informative to think about each of the quadrants of this plot. The upper right quadrant includes studies where the program improved attendance and delinquent behaviors. The lower left quadrant includes studies where the intervention made both behaviors worse relative

TABLE 3 Unstandardized intercept (B_0) and slope (B_1) between delinquency and nondelinquency effect sizes.

	B_0	B_1	95% CI		r^2	τ^2	k	Unweighted B_1
			Lower	Upper				
Psychological								
Attitudes about delinquency/prosocial conduct	0.07	0.13	-0.13	0.39	50	0.043	29	0.21
Personality traits	0.10	0.30	-0.03	0.63	54	0.073	35	0.25
Self-esteem/self-concept	0.09	0.22	-0.11	0.54	85	0.138	33	0.20
Internalizing/anxiety/depression	0.09	0.15	-0.05	0.35	30	0.014	29	0.17
Substance use	0.01	0.43	0.20	0.65	28	0.008	36	0.46
Interpersonal								
Family functioning	0.14	0.43	0.05	0.82	44	0.041	31	0.29
Peer relations	0.15	0.02	-0.24	0.28	49	0.030	33	0.06
Social skills	0.19	0.13	-0.21	0.46	31	0.030	21	0.14
School-related								
Attendance	0.08	0.52	0.35	0.69	30	0.019	50	0.50
Nondelinquent behavior resulting in discipline	0.08	0.38	0.03	0.72	72	0.076	22	0.40
Attitudes about school/teachers	0.00	0.36	0.06	0.66	46	0.015	25	0.30
Dropping out/graduating	0.08	0.28	-0.05	0.62	72	0.073	28	0.54
Academic achievement	0.12	0.13	-0.18	0.45	39	0.030	32	0.28
Vocational								
Employment status	0.06	-0.04	-0.31	0.23	52	0.020	31	0.15

Note: Number of independent study samples across all analyses = 173.

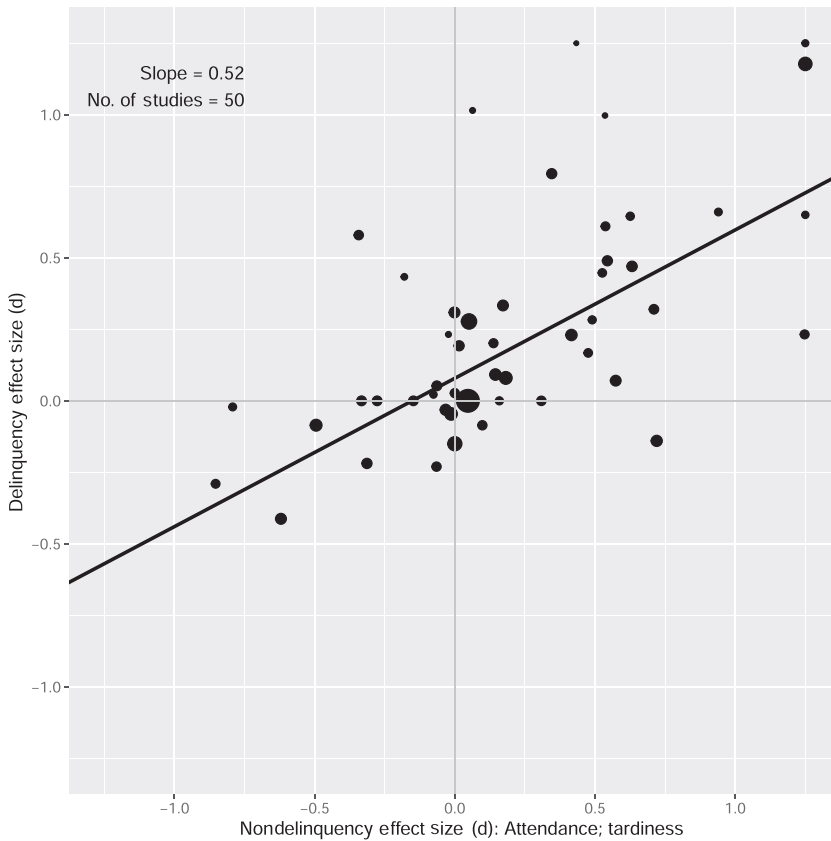


FIGURE 2 Scatter plot of delinquency effect sizes versus school attendance effect sizes.

to the comparison group. The relationships in both of these quadrants are consistent with this construct being a change lever. The upper left quadrant includes studies where delinquency was reduced, but attendance was not affected. Thus, these studies presumably reduced delinquency through some other change lever. The lower right contains studies that improved attendance but failed to improve delinquency. A high percentage of studies in this quadrant would raise doubts regarding the general robustness of this construct as a change lever. Figure 2 shows a clear pattern of evidence consistent with school attendance being a potential change lever. Most of the effects are in the upper right and lower left quadrants and only two of the 50 effects are in the problematic lower quadrant.

Along with the pattern of relatively strong delinquency relationships for several of the school-related constructs, there were notably strong relationships for the family functioning and substance use constructs ($B_1 = 0.43$ for both). Family-based treatment programs are reported to be among the more effective interventions for reducing delinquent behavior (Baldwin et al., 2012; Hartnett et al., 2017), and parenting approaches are strong predictors of delinquency (Hoeve et al., 2009). As such, we expected the family functioning construct to show a relatively strong relationship to delinquency effects. However, it is important to note that the interventions that produced these effects are not necessarily family therapy programs. Of the 31 studies represented, only 13 are evaluations of the effectiveness of family therapy of some form. Figure 3 shows the scatter plot for these effects and most of the studies are either in the upper right or lower left quadrants (21

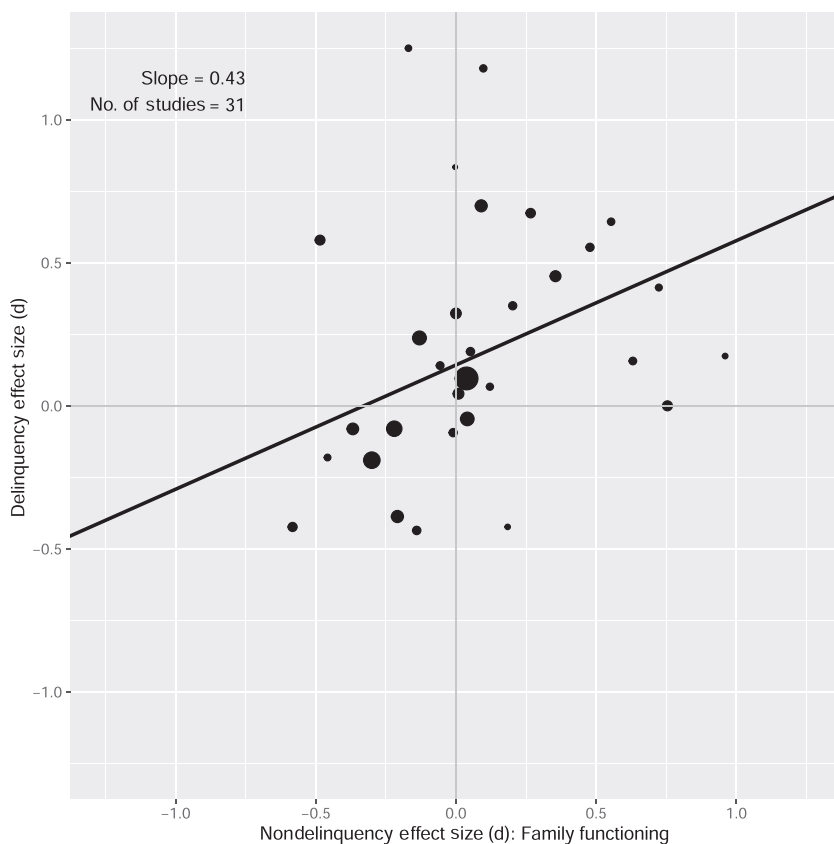


FIGURE 3 Scatter plot of delinquency effect sizes against family-functioning effect sizes.

of 31), as we would expect of an effective change lever. There are only two studies in the problematic lower right quadrant and one of these is near the origin of no effect on both constructs. Thus, this figure and the associated regression coefficient support the inference that family functioning may be an effective change lever.

The finding of a strong relationship between intervention effects on substance use and effects on delinquency ($B_1 = 0.43$) is also not surprising. Substance use has long been thought to both be causally related to the initiation of some forms of delinquency and to be caused by other forms of delinquency (Huizinga et al., 1989). It is thus quite plausible that reducing or preventing substance use could be a change lever for reducing subsequent delinquent behavior.

For two of the nondelinquency constructs, the regression coefficients showed essentially no relationship with delinquency effects. One of these was for employment status (e.g., getting or keeping a job) ($B_1 = -0.04$). Figure 4 shows the scatter plot for these data. In contrast to the scatter plot for school attendance, the effects are spread across all four quadrants, including numerous findings in the problematic lower right. Prior criminological research has suggested that employment for youth may actually increase delinquency (e.g., Ploeger, 1997), and the negative sign on this coefficient is consistent with that, though far from statistical significance. This finding raises doubts about the value of delinquency intervention programs focused on youth employment as the primary change mechanism. However, the effect may be quite different for young adults and we caution against over-generalizing this finding beyond school-age youth.

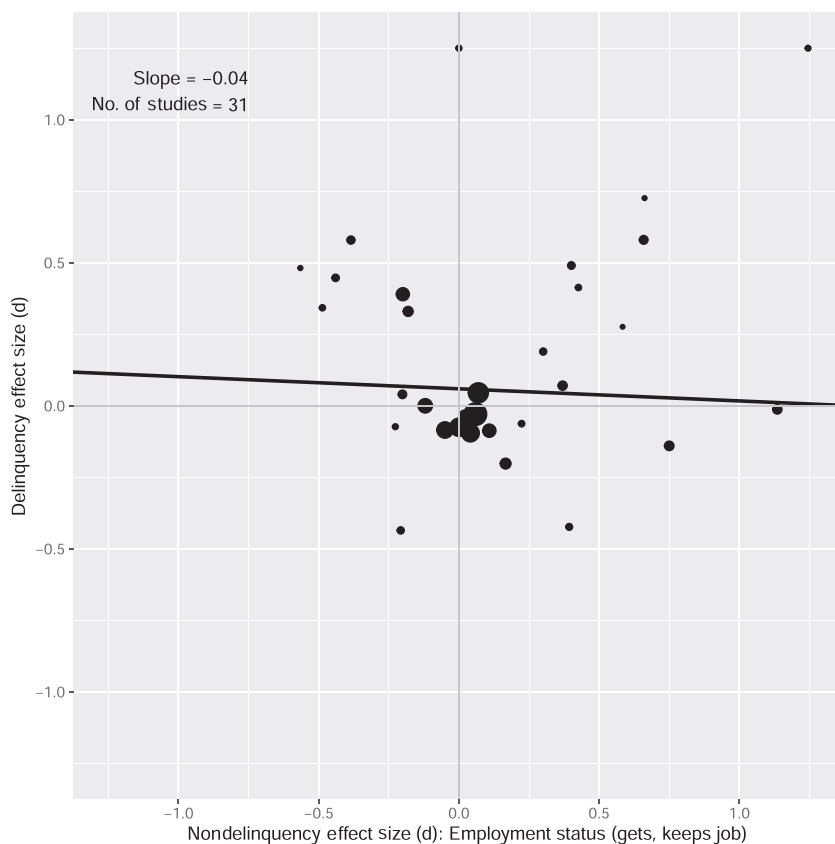


FIGURE 4 Scatter plot of delinquency effect sizes versus employment status effect sizes.

The weak finding for peer relations is counterintuitive ($B_1 = 0.02$). It is well established that youth are influenced by delinquent behavior among their peers (Gallupe et al., 2019). The outcome measures in this category are generally related to increasing association with prosocial peers, thus we would have expected a positive result for this construct. Additional research is needed to better understand this result and the role of altering peer relations in transitioning youth away from delinquency.

Small to moderate effects were observed for the remaining constructs of attitudes about delinquency/prosocial conduct ($B_1 = 0.13$), personality traits ($B_1 = 0.30$), self-esteem/self-concept ($B_1 = 0.22$), internalizing/anxiety/depression ($B_1 = 0.15$), and social skills ($B_1 = 0.13$). The largest of these, personality traits, involves such a mix of measures that it is hard to draw any meaningful conclusions regarding potential change levers from this finding. The relatively large coefficient, however, suggests that further research that provides more data for variables within this category might be productive.

2.2 | Sensitivity analyses

As sensitivity analyses, we used two alternative analytic approaches. The first was simply to compute the unweighted regression coefficients between the delinquency and nondelinquency effect

sizes. This analysis eliminates the differential influence of large versus small sample sizes, and the resulting coefficients are shown in the last column of Table 3. They track fairly closely to the inverse variance weighted estimates with a few meaningful differences. The estimate for the relationship of family functioning effects with delinquency is attenuated from 0.43 to 0.29 without the influence of several larger studies (see Figure 3). Among the school-related constructs, the estimated relationship for dropping out/graduating is much larger (increasing from 0.28 to 0.54), as is the relationship for academic achievement (increasing from 0.13 to 0.28).

The second sensitivity analysis addressed the possible differential influence of selected methodological features on the regression coefficients for the relationships between the nondelinquency and delinquency effect sizes. As described in the methods section, the delinquency effect sizes were adjusted for (a) whether a study was randomized, (b) whether the delinquency measure was based on official records, and (c) the log of the number of weeks posttest captured by the delinquency measure. The nondelinquency effect sizes were adjusted for (a) whether the study was randomized, (b) whether the outcome variable was a standardized measure, and (c) whether the measure was based on the youth's self-report. The results from these analyses are shown in Table 4. For most of the nondelinquency constructs, the effect of the method adjustments was to reduce the estimate of their relationship to delinquency, though nearly all the differences were modest. Overall, these findings track fairly closely to those from original analysis and do not alter the main conclusions. The largest differences were for social skills, with the coefficient dropping from 0.13 to a null -0.01 , attitudes about school/teachers (down from 0.36 to 0.26), and substance use (down from 0.43 to 0.34).

2.3 | Relationships between the nondelinquency constructs

It is unlikely that a single change lever is responsible for the delinquency effects found for effective interventions. Rather, we expect that multiple change levers are likely at play. We cannot assess the nature and effect of combinations of change levers with the available data. However, we can get a sense of the extent to which effects on one change lever tend to be associated with effects on other change levers by examining the correlations among effect sizes for the different nondelinquency constructs when they are assessed within the same studies. These correlations are shown in Table 5 for any pairing of constructs occurring in 10 or more studies. The upper triangle of this matrix shows the number of studies for each pairing, and the lower triangle shows the unweighted bivariate correlations.

Some of these correlations are quite high. For example, the correlation between the 16 pairs of effect sizes for attitudes about school/teachers and attitudes about delinquency/prosocial conduct is 0.81. Indeed, where sufficient data are available for an estimate, the school-related constructs are notable for generally high correlations with each other and with other constructs, especially in the psychological category. On the other hand, some correlations were notably low. One example is the 0.11 correlation across 11 studies for attitudes about delinquency/prosocial conduct and social skills. Despite the high correlations for some of these pairs, their change-lever effects were meaningfully different as, for example, with attitudes about school/teachers and attitudes about delinquency/prosocial conduct (see Tables 3 and 4). We caution, however, that the small sample sizes for these correlations make the results tentative, and their primary value is only to illustrate that intervention effects on change levers likely occur for multiple change levers at the same time rather than via single unique change pathways.

A further indication that delinquency interventions tend to influence multiple change levers rather than operating on a single primary pathway comes from an examination of the change

TABLE 4 Covariate-adjusted unstandardized regression coefficient between delinquency and nondelinquency effect sizes.

	B_0	B_1	95% CI		F^2	τ^2	k
			Lower	Upper			
Psychological							
Attitudes about delinquency, prosocial conduct	0.14	0.14	-0.11	0.38	36	0.024	29
Personality traits	0.15	0.24	-0.09	0.57	56	0.079	35
Self-esteem/self-concept	0.17	0.19	-0.14	0.52	84	0.129	33
Internalizing/anxiety/depression	0.12	0.16	-0.06	0.37	43	0.024	29
Substance use	0.08	0.34	0.10	0.57	38	0.012	36
Interpersonal							
Family functioning	0.24	0.41	0.04	0.79	31	0.023	31
Peer relations	0.20	-0.09	-0.35	0.17	46	0.027	33
Social skills	0.25	-0.01	-0.37	0.36	41	0.046	21
School-related							
Attendance	0.17	0.49	0.32	0.67	34	0.024	50
Nondelinquent behavior resulting in discipline	0.24	0.30	-0.07	0.68	75	0.091	22
Attitudes about school/teachers	0.09	0.26	-0.06	0.57	53	0.020	25
Dropping out/graduating	0.14	0.23	-0.10	0.56	69	0.064	28
Academic achievement	0.16	0.05	-0.26	0.35	45	0.038	32
Vocational							
Employment status	0.12	-0.04	-0.33	0.25	64	0.035	31

Note: These models are based on residualized effect sizes adjusted for methods covariates. For the delinquency effect sizes, the covariates were whether the study was random, whether the measure was based on official records, and the log of the weeks posttreatment captured by the delinquency measures. The nondelinquency effect sizes were adjusted for whether the study was random, whether the measure was a standardized instrument, and whether the measure was based on the youths' self-report. The residualized delinquency effect sizes were centered around the mean of a randomized study with an official measure measured at 52 weeks posttreatment. This centering only affects the intercept of the above models.

^aNumber of independent study samples across all analyses = 173.

TABLE 5 Unweighted correlation matrix among nondelinquency effect sizes (correlations in the lower triangle and number of pairwise comparisons in the upper triangle).

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Attitudes about delinquency/prosocial conduct		16	13	13	8	8	10	11	10	10	9	3	8	4
2. Self-esteem/self-concept	0.08		14	11	9	6	12	7	13	11	11	3	10	7
3. Personality traits	0.58	0.47		14	6	5	11	10	11	9	10	4	8	5
4. Internalizing/anxiety/depression	0.68	0.41	0.86		10	5	9	10	10	10	10	4	7	6
5. Substance use			0.62				4	11	8	8	8	5	7	5
6. Family functioning							14	6	7	3	3	2	5	5
7. Peer relations	0.49	0.47	-0.14		0.68	0.45		10	13	6	7	2	9	4
8. Social skills	0.11		0.36	0.29			0.27		6	6	6	1	5	3
9. Attitudes about school/teachers	0.57	0.81	0.83	0.74			0.71		11	11	9	4	9	4
10. Nondelinquent behavior resulting in discipline	0.57	0.64		0.48				0.43			17	3	14	2
11. Attendance		0.03	0.83	0.61					0.48			15	27	5
12. Dropping out/graduating											-0.31	9	10	
13. Academic achievement		0.72								0.59	0.07			2
14. Employment status												0.49		

Note: Only correlations for pairwise comparisons with a sample size of 10 or more are shown.

pathways affected by common types of interventions. Table 6 presents a selection of familiar types of interventions and reports the number of studies of each with both positive delinquency effects and positive effects on at least one change lever construct (effect sizes ≥ 0.10). The respective change-lever constructs for each intervention type are then listed. As is evident, the interventions that produce at least minimal positive delinquency effects also produce corresponding effects on a diverse range of change-lever constructs. We caution that the change-lever constructs identified from this perspective are heavily influenced by the choices made by the researchers about what to measure and the number of studies represented is relatively small. We do not believe this display provides a generalizable indication of the particular change levers or their range that is associated with the delinquency effects of any of the specific intervention types. However, it does illustrate the likelihood that multiple change pathways are typically involved in effective delinquency interventions.

3 | DISCUSSION

The large number of programs developed for addressing juvenile delinquency and the thin evidence base evaluating most of them presents a challenge for any policy maker or practitioner who wants to adopt evidence-based approaches. The implication of the meta-analytic evidence presented here is that a focus on change levers provides an alternative, and potentially more useful and parsimonious way of characterizing the likelihood that an intervention adopted in practice will be successful in reducing the recidivism of juvenile offenders. The logic of virtually all rehabilitation-oriented programs is to facilitate change in relatively proximal psychological, social, or behavioral propensities that, in turn, are expected to affect the likelihood of later delinquent behavior. There is only a modest number of intermediate outcomes of this sort that are plausibly associated with favorable effects on recidivism, and a variety of different programs may share a focus on one or more of them as the basis for their effectiveness.

Shifting the focus from lists of recommended programs to a categorization of programs according to the change levers they affect could facilitate the evolution of local programs toward more effective practice in various ways. Simply prioritizing programs that address effective change levers and avoiding those that concentrate on ineffective ones has the potential to increase the likelihood that the repertoire of programs a local agency uses will reduce recidivism. Awareness of effective change levers can also assist with matching offenders to appropriate programs. The assessment of the problem areas that contribute to a youth's delinquency, what is generally referred to as criminogenic needs, typically yields an identification of relevant proximal outcomes to target with selected interventions, for example, family issues, substance use, school engagement, and the like. An evidence base that identifies the outcomes of this sort that are demonstrably associated with reduced delinquency along with the types of programs that influence those outcomes would be a valuable tool for selecting programs likely to be effective with the respective juveniles.

Many juvenile justice agencies use structured risk/needs inventories to assess criminogenic needs, for example, the YLS/CMI, YASI, or PACT. In addition to identifying change lever outcomes appropriate to target with program services, these instruments can provide readily accessible data for determining if the provided services actually modify relevant change lever outcomes. Administered before and after service delivery, the extent of improvement on the change lever outcomes can be determined. For change levers known to be associated with reduced recidivism, these risk/need change scores thus provide quality control monitoring of the effectiveness of the respective services (Baglivio et al., 2017). Moreover, this assessment can be applied to any program in use

TABLE 6 Examples of nondelinquency outcomes with positive effects in at least one study of common interventions with positive delinquency effects.

Intervention	Nondelinquency outcome
Cognitive behavioral therapy (<i>k</i> = 5)	Internalizing/anxiety/depression Substance use Peer relations Social skills Attitudes about school/teachers Academic achievement
Family therapy (<i>k</i> = 10)	Self-esteem/self-concept Internalizing/anxiety/depression Family functioning Peer relations Social skills School attendance School nondelinquent behavior resulting in discipline Academic achievement
Group counseling (<i>k</i> = 7)	Attitudes about delinquency/prosocial conduct Personality traits Self-esteem/self-concept Family functioning Social skills School attendance School nondelinquent behavior resulting in discipline Attitudes about school/teachers Dropping out/graduating Academic achievement
Mentoring (<i>k</i> = 6)	Personality traits School attendance Dropping out/graduating Academic achievement Employment status
Social skills training (<i>k</i> = 4)	Self-esteem/self-concept Substance use Family functioning Peer relations Social skills Dropping out/graduating Employment status

Note: *k* = number of studies. The included studies have both a delinquency effect size and a nondelinquency effect size ≥ 0.10 .

by the agency to evaluate the likelihood of recidivism effects, including local programs that do not appear in any registry of evidence-based programs or which have been adapted to local contexts. In this way, evidence about change levers offers a basis for inferring a program's promise according to its ability to bring about change on one or more change levers related to reductions in delinquency.

Feedback from such monitoring of risk/needs change scores can also provide guidance for improving program implementation, the weak link in the transfer of promising program concepts to real-world adoption (Howell et al., 2014; Michie et al., 2013; Rhoades et al., 2012). In writing about psychotherapy, Kazdin (2007) argued that better knowledge regarding causal mechanisms, or what we call change levers, can facilitate implementation by helping frontline staff identify and support the critical elements of a program that are likely to be essential to maintaining effectiveness. In a similar vein, such knowledge can facilitate the development of effective new programs by concentrating attention on enhancements that increase the impact on effective change levers. It can also help assess the potential effectiveness of any newly proposed programs based on how well they map onto what we know works in terms of programmatic features and the change levers they act on.

Attaining any of these advantages for managing, improving, and developing effective programs for juvenile offenders requires a robust evidence base on the relationship of potential change levers to recidivism reduction. The meta-analysis results reported here fall well short of establishing that evidence base. The studies represented do not provide a systematic exploration of potential change levers but, rather, present data on a mix of intermediate outcomes of interest to the researchers in the specific contexts of the respective studies. Nor do any of them investigate the critical within-study relationships among study participants between intervention effects on the intermediate outcomes and those on later delinquency.

Despite those limitations, these meta-analysis results demonstrate the potential of the change-lever perspective and provide an initial sketch of some of its overall features. Foremost, those results show that some change levers are strongly enough associated with recidivism reduction to have practical significance. Among these were multiple school-related outcomes, including school engagement (i.e., improved attendance and reduced truancy), nondelinquent problem behaviors, and attitudes about school and teachers. In addition, effects on family functioning and substance use were also related to reductions in delinquent behavior. At the same time, these results reveal that not all the intermediate outcomes that researchers thought relevant to the interventions studied were, in fact, related to the delinquency effects of those interventions. Those less influential outcomes included getting/keeping a job, academic achievement, and peer relations.

Another notable finding about the potential change levers identified in the meta-analysis is that common intervention approaches often affect multiple different change levers and those effects are generally correlated. This is perhaps not surprising for programs such as family therapy, counseling, and mentoring that are designed to flexibly address any relevant family, school, and community issues. But it also appears for what are ostensibly more focused programs, for example, social skills training. Especially notable in this regard is the breadth of change levers potentially affected by programs that are not obviously oriented toward those outcomes. For example, of 31 studies reporting family functioning outcomes along with delinquency effects, only 13 of the interventions are family therapy. These findings suggest not only that there are multiple change levers affected by typical interventions but that those change levers are not necessarily ones explicitly targeted by the program design.

Building on these findings and their implications for effective practice, we believe that an ambitious research agenda is justified to establish a robust evidence base for change levers. We do not propose this as an alternative to further development of existing frameworks that focus on the

effectiveness of specific programs or the targeting of risk and protective factors for delinquency, but as an informative augmentation of those frameworks. The delinquency intervention literature is replete with theorizing about what makes programs effective and calls to get “inside the program black box.” But this has not yet led to development of generalized evidence regarding effective and ineffective change levers.

Most important for the development of such evidence is to greatly expanded the inclusion of direct tests of mediation at the level of the individual juvenile offenders represented in intervention studies. This can be done with readily accessible mediational analysis methods, such as structural equations, the Sobel mediation test (Baron & Kenny, 1986; Zhao et al., 2010), or a bootstrapping approach (Preacher & Hayes, 2008). All the studies that contributed to the meta-analysis presented here had the data necessary for such an analysis, yet none did so; a missed research opportunity. There are recent examples of these analyses in the juvenile delinquency treatment literature. Sheerin et al. (2021) used 10-year follow-up data from a randomized controlled trial of multisystemic therapy (MST) to investigate whether changes in family functioning, prosocial peer relations, and academic performance correlated with long-term reductions in violent, nonviolent, and drug-related crimes. While these analyses were correlational and restricted to the youth and families randomized to the MST condition, they did show that increased parental supervision was related to reductions in all crime types, and increased consistency in discipline was linked to reductions in nonviolent crime. In another example, Katz et al. (2022) tested for mediation using structural equation modeling methods in a randomized controlled trial of a family-based prevention program in Honduras. They found that improving family functioning, specifically by reducing family conflict and improving parental supervision, mediated the program’s effects on delinquency. They also found that increasing interactions with prosocial peers had an indirect, mediating impact on delinquency.

Analyses of mediation effects in individual studies would be most informative if a relatively broad range of plausible change lever constructs was included. In many studies this could be done without greatly encumbering the data collection, especially if data that might already be collected, or could easily be collected, from risk/needs assessment instruments were integrated into the design. An interesting distinction, however, is between interventions that focus primarily on a single potential change lever, for example, anger management programs, in comparison with those more broadly focused, such as counseling. A larger change may be produced with a narrow focus on a single change lever, possibly with correspondingly greater effects on delinquency, but effects on multiple change levers may have an aggregate effect that is advantageous even if the change in any given change lever is small. Where multiple change levers are affected, exploration of their interrelations and interactions would help clarify the nature of their influence on delinquency. Perhaps most important for a body of evidence about change levers, however, is simply identification of those with the largest effects on delinquency and the program characteristics that produce the largest effects on those change levers.

More generally, a broad benefit of a well-developed evidence base for change levers is the building blocks it would provide for program theory development. In their simplest form, program theories include three components: program activities (the active ingredients), some initial or proximal changes brought about by these activities (the potential change lever outcomes), and the longer-term practical or policy-relevant outcome (e.g., delinquency reduction). Program theory provides the rationale that connects these elements and the framework for improving existing programs and developing effective new ones. Unfortunately, very little empirical research currently examines these connections for programs treating juvenile offenders. The research agenda

we have sketched for change levers can contribute to developing such theory and its implications for program practice.

4 | CONCLUSION

We take an evolutionary and probabilistic view of evidence-based practice within juvenile justice. Henggeler and Schoenwald (2011) estimated that only 5% of high-risk offenders are treated with an evidence-based approach certified by either Blueprints or the Office of Justice Program's Model Programs registry (a precursor to the CrimeSolutions.gov registry). Elliott et al. (2020) have argued that the focus needs to be on getting more agencies to adopt programs shown to be effective and to do so with fidelity. In contrast, Lipsey (2020) contends that the "available evidence does not make a convincing case that these programs can be implemented and sustained in routine practice by juvenile justice personnel (i.e., without a 'high level of developer involvement') in a way that can be confidently expected to have a large overall impact on the recidivism outcomes of the juveniles served in a juvenile justice system" (p. 1333). We believe that a more realistic goal than universal adoption of programs from evidence-based registries is to find feasible ways to nudge a broad spectrum of juvenile justice practices towards approaches that successively increase the probability of better outcomes. Building on the widespread use of risk/needs assessment instruments, we believe that a well-developed evidence base that incorporates a change-lever perspective has the potential to do that.

CONFLICT OF INTEREST

The authors confirm that they have no conflict of interest to declare.

ORCID

David B. Wilson  <https://orcid.org/0000-0001-6461-3904>

REFERENCES

- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52(1), 7–27.
- Baglivio, M. T., Wolff, K. T., Jackowski, K., & Greenwald, M. A. (2017). A multilevel examination of risk/need change scores, community context, and successful reentry of committed juvenile offenders. *Youth Violence and Juvenile Justice*, 15(1), 38–61.
- Baldwin, S. A., Christian, S., Berkeljon, A., & Shadish, W. R. (2012). The effects of family therapies for adolescent delinquency and substance abuse: A meta-analysis. *Journal of Marital and Family Therapy*, 38(1), 281–304.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.
- Elliott, D. S., Buckley, P. R., Gottfredson, D. C., Hawkins, J. D., & Tolan, P. H. (2020). Evidence-based juvenile justice programs and practices: A critical review. *Criminology & Public Policy*, 19(4), 1305–1328.
- Gallupe, O., McLevey, J., & Brown, S. (2019). Selection and influence: A meta-analysis of the association between peer and personal offending. *Journal of Quantitative Criminology*, 35, 313–335.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34(4), 575–608.
- Gies, S. V., Nichols, L. M., Mojekwu, F., Guerette, R. T., & Tanner-Smith, E. E. (2023). Applying an empirically derived effect size distribution to benchmark the practical magnitude of interventions to reduce recidivism in the USA. *Journal of Experimental Criminology*, Advance online publication. <https://doi.org/10.1007/s11292-023-09561-1>
- Hartnett, D., Carr, A., Hamilton, E., & O'Reilly, G. (2017). The effectiveness of functional family therapy for adolescent behavioral and substance misuse problems: A meta-analysis. *Family Process*, 56(3), 607–619.

- Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, 117(1), 167.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65.
- Henggeler, S. W., & Schoenwald, S. K. (2011). Evidence-based interventions for juvenile offenders and juvenile justice policies that support them. social policy report. volume 25, number 1. *Society for Research in Child Development*.
- Hoeve, M., Dubas, J. S., Eichelsheim, V. I., Van der Laan, P. H., Smeenk, W., & Gerris, J. R. (2009). The relationship between parenting and delinquency: A meta-analysis. *Journal of Abnormal Child Psychology*, 37, 749–775.
- Howell, J. C., Lipsey, M. W., & Wilson, J. J. (2014). *A handbook for evidence-based juvenile justice systems*. Lexington Books.
- Huizinga, D. H., Menard, S., & Elliott, D. S. (1989). Delinquency and drug use: Temporal and developmental patterns. *Justice quarterly*, 6(3), 419–455.
- Katz, C. M., Cheon, H., Hedberg, E., & Decker, S. H. (2022). Impact of family-based secondary prevention programming on risk, resilience, and delinquency: A 6-month follow up within a randomized control trial in Honduras. *Justice Quarterly*, 39(6), 1237–1262.
- Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, 3, 1–27.
- Koehler, J. A., Lösel, F., Akoensi, T. D., & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *Journal of Experimental Criminology*, 9, 19–43. <https://doi.org/10.1007/s11292-012-9159-7>
- Lipsey, M. W. (1992a). The effect of treatment on juvenile delinquents: Results from meta-analysis. In (F. Losel, D. Bender, & T. Bliesener Eds.), *Psychology and law: International perspectives* (pp. 131–143). Walter de Gruyter.
- Lipsey, M. W. (1992b). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In (T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, & R. J. Light Eds.), *Meta-analysis for explanation: A casebook* (pp. 83–127). Russell Sage Foundation.
- Lipsey, M. W. (1995). What do we learn from 400 research studies on the effectiveness of treatment with juvenile delinquents? In (J. McGuire Ed.), *What works: Reducing reoffending: Guidelines from research and practice* (pp. 63–78). John Wiley & Sons.
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and offenders*, 4(2), 124–147.
- Lipsey, M. W. (2020). Revisited: Effective use of the large body of research on the effectiveness of programs for juvenile offenders and the failure of the model programs approach. *Criminology & Public Policy*, 19(4), 1329–1345.
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* [NCSEER 2013-3000]. National Center for Special Education Research, Institute for Educational Sciences, U.S. Department of Education.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.
- Mendoza, J. L., & Mumford, M. (1987). Corrections for attenuation and range restriction on the predictor. *Journal of Educational Statistics*, 12(3), 282–293.
- Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, 46(1), 81–95.
- Ploeger, M. (1997). Youth employment and delinquency: Reconsidering a problematic relationship. *Criminology*, 35(4), 659–676.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, 40(3), 879–891.
- Pustejovsky, J. E., & Chen, M. (2024). Equivalencies between ad hoc strategies and multivariate models for meta-analysis of dependent effect sizes. *Journal of Educational and Behavioral Statistics*. <https://doi.org/10.3102/10769986241232524>
- Rhoades, B. L., Bumbarger, B. K., & Moore, J. E. (2012). The role of a state-level prevention support system in promoting high-quality implementation and sustainability of evidence-based programs. *American Journal of Community Psychology*, 50(3–4), 386–401.

- Sanchez-Meca, J., Marin-Martinez, F., & Chacon-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8*(4), 448.
- Sheerin, K. M., Borduin, C. M., Brown, C. E., & Letourneau, E. J. (2021). An evaluation of mechanisms of change in multisystemic therapy for juvenile justice-involved youths a decade following treatment. *Journal of Marital and Family Therapy, 47*(1), 208–219.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*(2), 197–206.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wilson, D. B., & Lipsey, M. W. (2024). Scaling up effective juvenile delinquency programs by focusing on change levers: Evidence from a large meta-analysis. *Criminology & Public Policy, 23*, 261–286.
<https://doi.org/10.1111/1745-9133.12663>

AUTHOR BIOGRAPHIES

David B. Wilson is a distinguished university professor in Criminology, Law and Society at George Mason University. His research interests include the effectiveness of adult and juvenile treatment and crime prevention programs and policies, program evaluation methodology, meta-analysis, and statistical methods.

Mark W. Lipsey is a research professor in Human and Organizational Development at Vanderbilt University. His research specialties are evaluation research and meta-analysis investigating the effects of interventions with children, youth, and families.