The Impact of Teen Courts on Youth Outcomes: A Systematic Review

Lauren N. Gase^{1,2} · Taylor Schooley² · Amelia DeFosset² · Michael A. Stoll³ · Tony Kuo^{2,4}

Received: 22 June 2015 / Accepted: 19 July 2015 / Published online: 28 July 2015
© Springer International Publishing 2015

Abstract Processing juvenile offenders in the traditional justice system can lead to a range of negative consequences. As an alternative to formal criminal processing, many jurisdictions have begun to implement diversion programs for first-time or low-level offenders. This systematic review sought to summarize evidence of the effectiveness of one commonly used diversion model, Teen Courts, on outcomes for juvenile offenders. Teen Courts were defined as any intervention for youth (ages 10–17) in which the participating offenders' peers were involved in verdict or sentencing. Final analysis included 22 studies. Among the 15 studies that assessed statistical significance of recidivism, 4 found statistically significant results favoring Teen Courts, 1 found statistically significant results favoring the traditional justice system, and 10 found null results. Most studies provided little detail regarding the structure or approach of Teen Courts under study and varied widely in research design, comparison group, and operationalization of recidivism, making it

difficult to compare results. In order to inform decision-making about the use of Teen Courts, additional studies are needed that maximize internal and external validity, consider pathways of intervention effects, and examine potentially differential impacts of the program on participating youth.

Keywords Teen court · Youth court · Peer court · Diversion · Restorative justice · Juvenile justice

Introduction

Processing juvenile offenders in the traditional justice system can lead to a range of negative consequences. Confinement models, for example, can result in the removal of youth from the presence of positive, caring relationships with adults and peers, and from opportunities for community involvement that foster critical thinking, self-reliance, self-efficacy and other developmental assets, all basic conditions necessary for prosocial development (National Research Council 2013). In as much as formal juvenile justice systems tend to remove “problem” juveniles from families, peers, and communities, the psychosocial effect for youth is the likely interruption of developmental processes and the heightened potential for further system involvement. Formal processing can be deleterious even for youth who do not face incarceration. For example, court processing that results in a criminal record can stymie future opportunities (e.g., in employment and higher education) and establish a negative life trajectory, increasing the likelihood of incarceration amid a host of other poor academic, health, and behavioral outcomes (Nellis 2011). Because of these and other observations about youth outcomes, there have been calls to reform the U.S. juvenile justice system and develop diversionary

✉ Lauren N. Gase
lgase@ph.lacounty.gov

¹ Department of Health Policy and Management, University of California, Los Angeles Fielding School of Public Health, 650 Charles E. Young Dr. South, Los Angeles, CA 90095, USA

² Division of Chronic Disease and Injury Prevention, Los Angeles County Department of Public Health, 3530 Wilshire Blvd, 8th Floor, Los Angeles, CA 90010, USA

³ University of California, Los Angeles, Luskin School of Public Affairs, 3250 Public Affairs Building, Box 951656, Los Angeles, CA 90095, USA

⁴ Department of Epidemiology, University of California, Los Angeles Fielding School of Public Health, 650 Charles E. Young Dr. South, Box 951772, Los Angeles, CA 90095, USA

models that help youth avoid formal criminal processing, especially for first-time or low-level offenders (National Research Council 2013).

Teen Courts, also referred to as youth, peer, or student courts, represent one model of diversion that has been widely embraced across the U.S. Beginning as a local effort in the 1970s, Teen Courts were rapidly popularized in the 1990s, with the number of courts growing from 78 in 1994 to over 1000 in 2010. Teen Court programs now operate in 49 states and the District of Columbia (National Association of Youth Courts 2015), and see over 100,000 cases per year (American Bar Association 2007). Teen Courts commonly include a limited case pool of first-time and/or low-level offenders, volunteer youth taking an active role in providing consequences to the offender (e.g., as juror, attorney, or judge), and the use of future Teen Court jury service as a primary sanction. Additionally, many Teen Court programs offer sanctions directed at repairing harms caused by the offense, such as community service, letters of apology to victims or others, or essays (National Association of Youth Courts 2015). Although Teen Courts vary in their structure and approach (e.g., agencies involved in administration, funding sources, participation criteria for offenders) (Nessel 2000), their ultimate goal is to determine a fair and restorative sentence or disposition for the juvenile offender (National Association of Youth Courts 2015).

Although multiple studies have examined Teen Court programs, there is conflicting evidence as to whether and how these programs impact the behavior of participating offenders (e.g., reduce recidivism and prevent crime). For instance, Butts et al. (2002) evaluated outcomes across four geographically diverse Teen Court programs, of which two sites demonstrated significantly lower rates of recidivism after 6 months for youth processed through a Teen Court vs. traditional court; 1 site showed no difference; and the fourth demonstrated a slight, although non-significant, trend in favor of the comparison condition. Similar results suggesting potentially unfavorable outcomes associated with Teen Court program participation have been reported elsewhere (Povitsky 2005; Wilson et al. 2009). Such contradictory findings proliferate in the literature on Teen Courts. These mixed findings may result from the wide variation in the methodological consistency and quality of research studies conducted to date. For example, few effectiveness studies have included a comparison group and even fewer have included one that is methodologically appropriate (e.g., comparable on key variables such as age or type of offense) (Stickle et al. 2008; Administrative Office of the Courts 1995). While recent efforts to describe the current state of the evidence on Teen Courts provide a useful starting place to understand the impact of such programs, these studies have fallen short in their ability to systematically identify all relevant literature, examine direction and size of program

effects, and consider variations in methodological quality when synthesizing study results (Puzach and Haas 2014; Schwalbe et al. 2012). These considerations point to the need for additional work to create a comprehensive empirical synthesis of how and to what extent youth are impacted by participation in Teen Courts.

Current Study

To inform program planning and define directions for future research, this study sought to systematically identify and summarize evidence of the impact of Teen Courts on outcomes for juvenile offenders. This work was guided by the following research questions: (1) what are the essential characteristics of Teen Court programs examined in the literature?; (2) what study designs, methods, and outcome measures have been used to examine the effectiveness of Teen Court programs?; (3) what evidence exists regarding the impact of Teen Court programs on outcomes for juvenile offenders?; and (4) what are the strengths and weakness of the current evidence? To help contribute to a more nuanced understanding of the features of Teen Courts with the greatest opportunity for positive impact, we sought to identify predominant theoretical assumptions and program features and quantitatively synthesize and link these, where feasible, to evidence of potential short- and long-term outcomes. In recognition of the ongoing need to identify a full range of potential impacts and refine program theory, studies were included that examined any potential outcome for juvenile offenders who participated in Teen Court programs. In order to generate a comprehensive picture of the current state of the evidence base, we sought to systematically identify evidence from a broad range of sources, including those outside of traditional peer-reviewed venues such as reports, dissertations and theses, and take into account methodological strength when assessing and summarizing results.

Methods

This systematic review of the evidence was conducted in line with Institute of Medicine and Evidence for Policy and Practice standards. Protocol for independent search, screening, analysis, reporting, and managing potential bias were outlined by the research team (authors and reviewers) at the outset of review activities (Eden et al. 2011; Evidence for Policy and Practice 2007).

Search Strategy and Analysis Sample

The search strategy, including scope of included databases and search parameters and terms, was developed in

consultation with biomedical research librarians from the University of California, Los Angeles, as well as content experts with knowledge of criminal justice research and mixed methods systematic reviews. Studies were identified through an electronic search of the National Criminal Justice Review Service (NCJRS), Science Direct, Web of Science, PsychINFO, Google Scholar, Urban Institute, ProQuest Digital Dissertations & Theses, *Criminology*, and the *American Journal of Criminal Justice*. These sources were chosen for their multidisciplinary nature and relevance to research on Teen Courts.

The initial search occurred between October 2014 and January 2015, using the search terms (“teen court” OR “youth court” OR “peer court” OR “student court”) AND (“evaluation” OR “impact” OR “assessment” OR “outcome”). In searching both ProQuest Digital Dissertations and Google Scholar, “teen court” AND (“evaluation” OR “impact” OR “assessment” OR “outcome”) was used; terms were modified in order to balance the sensitivity and specificity of the search in accordance with Evidence for Policy and Practice standards (2007) (Gough et al. 2012).

The search identified a total of 1929 sources. Titles and abstracts for all sources were initially reviewed by one member of the research team (reviewer). Sources were included for the next stage of full-text review if they: (a) were published in English after 1980, (b) examined programs that aligned with the review’s definition of Teen Courts, (c) involved an experimental or quasi-experimental research design, and (d) involved analysis of primary or secondary data to assess any outcome of participation in Teen Courts on juvenile offenders.

For the purposes of this review, Teen Courts were defined as a diversion program for juvenile offenders between the ages of 10 and 17 in which youth were involved in the verdict and/or sentencing of their peers (Administrative Office of the Courts 2011). Reviewers made no exclusions based on geographic location or court setting. Of note, all international sources identified during the initial review were excluded because none, despite their use of the term teen, youth, peer, or student court, described programs that aligned with the review’s definition of Teen Courts. Each source identified in the search was required to assess at least one outcome (e.g., knowledge, attitudes, and behaviors) for juvenile offenders who participated in Teen Court programs. Sources that explored the impact of Teen Courts on volunteer youth (e.g., peer jurors) who were not current offenders were excluded from the review. The review included only sources that used experimental or quasi-experimental designs. Although observational studies can be useful in understanding program functioning, their lack of a control or comparison condition weakens internal validity (Campbell and Stanley 1966). To facilitate synthesis of the results of studies with

strong internal validity, we did not include observational studies in our final analysis sample.

Sample Selection

Based on the review criteria, 73 documents (articles, reports, and dissertations/theses) of the 1929 sources were identified for independent full-text review; these 73 documents were reviewed by two members of the research team (reviewers) to determine final eligibility. Each reviewer made recommendations for inclusion or exclusion independently of one another. They then met to discuss recommendations and finalize the analysis sample. Studies described in the documents were excluded from the final sample, in accordance with the review’s defined inclusion criteria, if they did not: (a) align with the review’s definition of Teen Courts (24 excluded), (b) use an experimental or quasi-experimental design (19 excluded), (c) examine outcomes for juvenile offenders (7 excluded), or (d) report primary data (4 excluded) (Fig. 1). All references cited in the 19 eligible documents were also screened for eligibility, yielding two additional documents that met inclusion criteria. In total, 21 documents met all criteria and were included in the final sample.

Data Abstraction

Data abstraction was completed independently by two members of the research team (reviewers) between January and March 2015 using a standardized abstraction tool. Documents were coded in batches of 3–6. After each batch, reviewers met to discuss results and address discrepancies. Data elements were abstracted for each “study,” defined as an individual Teen Court or group of courts (Teen Court program) within the same jurisdiction in which the results were pooled. Documents which reported results of the same court/program during the same time period were coded together (as the same “study”).

The data abstraction tool was developed iteratively by the two reviewers who carried out the abstraction procedures. The tool included 22 items in 5 sections: (1) basic information (year, type of publication); (2) program characteristics (number of court sites/geographic location, program partners, court model, program length/requirements, participation criteria, how program participants were identified); (3) study design (research design, sample, subject demographics, participation/response rates, comparison group, outcomes, measurement, statistical analyses); (4) results (main effects, sub-group analyses/effects, sensitivity analyses/effects); and (5) strengths and weaknesses (internal validity, external validity, fidelity and quality of intervention).

Fig. 1 Summary of systematic review of Teen Courts, flow diagram

Basic Information

Reviewers noted the year of publication and the type of publication (coded as peer-reviewed article, report, or dissertation/thesis) for each source.

Program Characteristics

Reviewers coded the number of court sites and geographic location (city, region, state) examined in each study. Partners were coded by noting any organizations involved in administering the Teen Court program or providing financial or in-kind support. The court model was coded according to the judicial field's four established models of Teen Courts: adult judge, youth judge, youth tribunal, and peer jury (Butts et al. 2002). When the article noted a specific program model, that model was noted; when a

model was not named, reviewers coded the court model by comparing program descriptions to established definitions. To code program length and requirements, reviewers noted any information about the duration of the program, any protocol for referral to Teen Courts, types of sentencing provided, and protocols for facilitating monitoring and compliance. Participation criteria were coded by noting any information on type(s) of youth (e.g., age, prior offense) and types of offense accepted. Reviewers also abstracted information on the methods by which a Teen Court's staff identified or referred eligible participants.

Study Design

Reviewers coded research design according to the criteria defined by Campbell and Stanley (1966). To code study

sample, reviewers abstracted any information about sample criteria (e.g., court years, program status) as well as sample size. Reviewers noted any information on participation rate (percent who agreed to participate in the study), response rate (percent who responded to the survey or for whom data could be abstracted), and attrition rate (percent who dropped out of the study). Reviewers noted sample demographics, including gender, average or age range, and race/ethnicity. To code comparison groups, reviewers abstracted any information on the composition of each group (e.g., geographic location, type of offense, program years), how they were selected (e.g., matching criteria), and sample size. To code outcomes, reviewers noted how each outcome construct was operationalized (e.g., type of outcome, time frame for outcome) as well as how it was measured (e.g., administrative records, qualitative interviews). All statistical analyses used to assess outcomes were also noted.

Results

Reviewers abstracted results for each outcome reported, including magnitude of the results as well as statistical significance. Any results of sensitivity analyses (e.g., different statistical models, exclusion of uncertain data points) were abstracted. In addition, any results from the analysis of sub-groups (either differential impact of the program on types of youth or impact of specific program components) were noted.

Strengths and Weaknesses

Reviewers noted strengths and weaknesses of the study in terms of internal validity, external validity, fidelity, and quality of the intervention. When assessing internal validity, reviewers considered the effects of history or time between measurements, maturation, testing, instrumentation, statistical regression, bias of sample selection, differential loss to follow up, and other possibilities for confounding between intervention and comparison groups (Campbell and Stanley 1966; Valentine and Cooper 2008). When assessing external validity, reviewers considered reaction to measurement, appropriate follow-up, intervention dose and characteristics, and relevance of the sample to the larger community (Campbell and Stanley 1966; Valentine and Cooper 2008). When assessing fidelity and quality of the intervention, reviewers considered the extent to which the intervention reflected commonly held or theoretically derived characteristics (e.g., principles of restorative justice), whether the intervention was described at a level of detail that would allow its replication by others, the relevance of outcomes measured, and the appropriateness of follow-up time (Valentine and Cooper 2008).

Results

Review of the 21 documents yielded analyses for 22 studies of an individual Teen Court or Teen Court program. Studies were published in non-peer-reviewed reports or working papers ($n = 6$), peer-reviewed publications ($n = 11$) and dissertations or theses ($n = 4$). Years of publication (when dates were provided) ranged from 1987 to 2014, with the majority (13 documents) published during or after 2005. After review of the studies, it was determined that the scarcity of methodologically rigorous studies and wide variation in program characteristics (e.g., types of offenders), study design (e.g., composition of the comparison group) and outcome measures (e.g., definition of recidivism, length of study follow-up) in the current literature on Teen Courts would limit the ability of meta-analysis to present meaningful aggregative synthesis (Valentine and Cooper 2008; Gough et al. 2012). This review therefore presents a systematic descriptive analysis and sets the stage for future meta-analyses by providing recommendations for enhancing the quality of future research and evaluation of Teen Courts.

Teen Court Program Characteristics

Overall, most studies in the analysis sample provided few details on the specific structures and approaches used by the Teen Courts under study. Teen Courts differed in criteria for offender participation: 17 studies assessed courts that accepted most low-level misdemeanor offenses (e.g., theft/burglary, assault, drug/alcohol possession), 4 of these courts excluded any violent offenses and 3 excluded all or certain drug offenses. Four studies examined courts that accepted somewhat distinct offense types: only alcohol or tobacco possession (Patrick and Marsh 2005), mostly vehicle-related offenses (Hissong 1991), and school-based offenses (Administrative Office of the Courts 1995; Norton et al. 2013). The most common age criteria (reported by 9 studies) required youth to be between the ages of 11 or 12 and 17 or 18. Only 4 studies examined courts that allowed younger offenders (as young as 7). Most commonly, studies assessed courts which required youth to be first time offenders (4 of the 13 with this requirement made exceptions on a case-by-case basis); fewer (3 studies) examined courts that allowed repeat offenders (Table 1).

Teen Courts differed in their court model and processes. Of the 20 studies for which a court model was described, 7 used an adult judge model, 5 used a youth judge model, 3 used mixed models (adult judge/peer jury), 1 used a youth tribunal model, and 4 used a peer jury model. Nine studies mentioned that youth were required to admit guilt prior to participation in the Teen Courts' hearing (i.e., the hearing

Table 1 Program characteristics of studies included in the systematic review of Teen Courts, 2015

Author (year)	Court model	Admission of guilt required?	Types of charges accepted into teen courts	First time offenders?	Ages accepted	Program length
<i>Randomized experimental designs</i>						
Patrick et al. (2004); Patrick and Marsh (2005)	Adult judge	?	Only tobacco or alcohol possession	Yes	12–18	?
Stickler et al. (2008); Dugas (2006); Wilson et al. (2009)	?	Yes	Various misdemeanor offenses	Yes	11–17	?
<i>Quasi-experimental non-equivalent control group designs</i>						
Nochajski et al. (n.d.)	Youth judge	Yes	Various nonviolent misdemeanor offenses	Yes	7–16	Average 7 weeks
Seyfrit et al. (1987)	Peer jury	Yes	Nonviolent offenses; alcohol- or drug-related, burglary, shop-lifting, and theft most common	?	11–17	Up to 24 weeks
Hissong (1991)	Adult judge	Yes	Class C misdemeanor offenses; vehicular offenses most common	?	?	?
Administrative Office of the Courts (1995)	Adult judge	Yes	School-based misbehavior or nonviolent offenses that would be misdemeanors or infractions if committed by an adult; property offenses most common	Most	10+	8 weeks
Logalbo and Callahan (2001)	Adult judge	?	?	?	?	5 weeks
Butts et al. (2002) (Anchorage)	Youth tribunal	No	Minor property and misdemeanor (e.g. shoplifting, trespassing, vandalism); no drug- or alcohol-related offenses	Most	10+	?
Butts et al. (2002) (Tempe and Chandler)	Mix (adult judge/peer jury)	?	Minor civil and criminal offenses, including traffic and local ordinance violations; shop-lifting most common	?	12–17	?
Butts et al. (2002) (Montgomery County)	Mix (adult judge/peer jury)	?	Shop-lifting most common, also referenced vandalism and alcohol-related offenses	Most	12–17	?
Butts et al. (2002) (Independence)	Youth judge	No	Status offenses (e.g. truancy, curfew, underage drinking), third degree assault, minor drug- and alcohol-related offenses; shop-lifting, truancy, and vandalism most common	No	7–16	?
Weisz et al. (2002)	Adult judge	?	Shop-lifting most common	?	?	16 weeks
Povitsky (2005)	?	Yes	Misdemeanor offenses; no drug-related offenses more serious than marijuana use, possession, or paraphernalia	Yes	11–18	?
Forgays (2008); Forgays and DeMilio (2005)	Adult judge	Yes	Misdemeanor or gross misdemeanor offenses; theft and/or shop-lifting, assault, alcohol or marijuana possession most common	No	?	12 weeks
Norris et al. (2011)	Adult judge	Yes	Status or low-level misdemeanor offenses; no drug-related or violent/serious offenses	Yes	?	12–16 weeks

Table 1 continued

Author (year)	Court model	Admission of guilt required?	Types of charges accepted into teen courts	First time offenders?	Ages accepted	Program length
Bright et al. (2013) (Baltimore City)	Mix (adult judge/peer jury)	?	Misdemeanor offenses	No	11–17	?
Bright et al. (2013) (Charles County)	Peer jury	?	Theft, assault, and traffic violations most common	Yes	12–17	?
Bright et al. (2013) (Montgomery County)	Peer jury	?	Theft, alcohol-related offenses, property offenses, fourth degree burglary, and marijuana possession most common	Most	12–17	?
Jacobsen (2013)	Youth judge	?	Misdemeanor or gross misdemeanor offenses	Yes	?	?
Laundra et al. (2013)	Peer jury	?	Theft, drug-related offenses, and simple battery most common	Yes	?	?
Norton et al. (2013)	Youth judge	Yes	School-based low-level infractions	?	?	2–4 weeks
Buchholz (2014)	Youth judge	?	?	Yes	?	?

? = Unable to be determined based on document(s) reviewed

was focused primarily on understanding youth characteristics, needs and appropriate sentences), whereas 2 studies described an assessment and determination of guilt or innocence during the Teen Court hearing. Types of sentences required were fairly consistent across the courts under study, including community service, letters of apology, and serving on a jury in future Teen Court sessions. For the 8 studies that mentioned the average or maximum length of time an offender spent in the Teen Court program, time ranged from 2 to 24 weeks (Table 1).

Few studies described the theories on which Teen Courts under study were based. Five studies mentioned courts being guided by restorative justice principles (e.g., focused on the reparation of harms between the offender and the victim or community). No studies included a logic framework or logic model that described how intervention components were expected to drive outcomes. In fact, Norton et al.'s (2013) evaluation of school-based courts in Pennsylvania included explicit recommendations on the need to create program theory about the ways in which Teen Courts could impact the outcomes under study (subsequent disciplinary infractions). Only two studies (Stickle et al. 2008; Forgays and DeMilio 2005) described having conducted assessments to examine the fidelity of the courts to a theoretical model.

Study Design and Samples

Twenty of the 22 studies used a non-equivalent control group quasi-experimental design. Only 2 studies employed random assignment using a post-test only design. One of the randomized studies used an intent-to-treat design whereas the other included only youth who completed the program in the analysis sample; no details were provided on completion criteria, rate, or reasons for not completing the program.

Among the quasi-experimental studies, 11 selected juvenile offenders (that did not participate in the Teen Court program) during the same time period as the study as the comparison group; most obtained data on youth who lived within the jurisdiction under study but 3 obtained data on youth from another (neighboring) jurisdiction. Three studies used data from a historical comparison group (before the Teen Court program was implemented/scaled in their jurisdiction). Four studies selected youth who were referred to the program but did not participate as the comparison group and two studies selected youth who were not involved in the justice system as the comparison group (Table 2). Most quasi-experimental studies established inclusion criteria and/or used matching in an effort to make the comparison group comparable to youth who participated in the Teen Court program (i.e., meeting age and offense criteria that would have qualified them for a Teen

Table 2 Study characteristics and results of studies included in the systematic review of Teen Courts, 2015

Author (year)	Intervention sample	Control/comparison group	Comparison condition/program	Outcomes measured	Way recidivism measured	Measurement time frame	Analyses	Main effects	Sub-analyses
<i>Randomized experimental designs</i>									
Patrick et al. (2004); Patrick and Marsh (2005)	Program completers (n = 75)	Randomized offenders (n = 68)	Diversion program, brief intervention, or traditional justice processing	Recidivism	Citation/arrest	3 years	Chi square	43 % recidivism in TC vs. 38 % in juvenile accountability, 43 % in magistrate court, and 50 % in educational control (n.s.)	
Stickle et al. (2008); Dugas (2006); Wilson et al. (2009)	Program referrals (n = 56)	Randomized offenders (n = 51)	Traditional justice processing	Attitudes/beliefs; behaviors; recidivism	Citation/arrest	18 months	Chi square	TC group had higher frequency of delinquent behavior ($p = .024$); no differences in drug use, social skills, beliefs, self-concept, rebelliousness, or neighborhood attachment; 32 % recidivism in TC vs. 26 % in control (n.s.)	No significant interaction between program and gender on self-concept, last month frequency drug use, variety drug use or delinquent behavior
<i>Quasi-experimental non-equivalent control group designs</i>									
Nochajski et al. (n.d.)	Program participants (n = 55)	Other offenders (n = 65)	Diversion program	Recidivism; residential placement	Citation/arrest	6 months; 1 year; varied	Survival analysis; Chi square	No significant difference between the two groups in recidivism; 3.6 % of TC youth placed in residential setting vs. 10.8 % of comparison youth ($p = .139$)	Workshop attendance showed a significant effect on combined recidivism (odds ratio = 23.58, $p < .001$)
Seyfrit et al. (1987)	Program participants (n = 52)	Other offenders (other jurisdiction) (n = 50)	Diversion program	Recidivism	Referral	Varied: 9–27 months	t test	9.6 % of TC participants committed an additional offense compared with 12 % in comparison group ($p > .10$)	
Hissong (1991)	Program participants (n = 196)	Other offenders (n = 196)	Traditional justice processing	Recidivism	?	Varied: up to 2 years	Survival analysis	Probability of surviving beyond 18 months for TC participants was 74.7 % compared with 64.1 % for comparison group ($p < .05$)	Significant differences in survival curves for males; no significant differences in survival curves for females
Administrative Office of the Courts (1995)	Program referrals (n = 95)	Historic (n = 97)	Diversion program	Recidivism	Referral	7.4 months	Multivariable regression	No significant relationship between program and recidivism beyond that accounted for by differences in age and offense	

Table 2 continued

Author (year)	Intervention sample	Control/comparison group	Comparison condition/program	Outcomes measured	Way recidivism measured	Measurement time frame	Analyses	Main effects	Sub-analyses
Logalbo and Callahan (2001)	Program completers (n = 111)	Non offenders (n = 65)	None	Knowledge and attitudes	–	5 weeks	Multivariable regression	Increase in knowledge ($p < .05$); increase in positive attitudes toward judge ($p < .05$); decrease in positive attitudes toward police ($p < .05$) among TC program completers compared to high and middle school students	
Butts et al. (2002) (Anchorage)	Program referrals (n = 120)	Historic (n = 120)	Traditional justice processing	Recidivism	Referral	6 months	Chi square	6 % recidivism in TC vs. 23 % in comparison ($p < .01$)	
Butts et al. (2002) (Tempe and Chandler)	Program referrals (n = 115)	Other offenders (n = 115)	Traditional justice processing	Recidivism	Referral	6 months	Chi square	9 % recidivism in TC vs. 15 % in comparison (n.s.)	8 % recidivism in adult judge court vs. 10 % in peer jury court (n.s.)
Butts et al. (2002) (Montgomery County)	Program referrals (n = 154)	Other offenders (other jurisdiction) (n = 62)	Diversion program	Recidivism	Citation/arrest	6 months	Chi square	8 % recidivism in TC vs. 4 % in comparison (n.s.)	12 % recidivism in adult judge court vs. 5 % in peer jury court (n.s.)
Butts et al. (2002) (Independence)	Program referrals (n = 142)	Other offenders (n = 142)	Traditional justice processing	Recidivism	Referral	6 months	Chi square	9 % recidivism in TC vs. 28 % in comparison ($p < .01$)	
Weisz et al. (2002)	Program referrals (n = 177)	Non offenders (n = 175)	None	Attitudes; recidivism	Referral	1 year	Descriptive; paired <i>t</i> test	No significant change in TC participant attitudes over time in comparison to jury participants or high school students; 13 % of TC referrals had re-offended within a year compared with a 24 % re-offense rate for the County's entire diversion program	
Povitsky (2005)	Program participants (n = 211)	Other offenders (other jurisdiction) (n = 781)	Traditional justice processing	Recidivism	Citation/arrest	18 months	Multivariable regression	Odds ratio of reoffending (TC vs. comparison) = 2.297 ($p < .05$)	
Forgays (2008); Forgays and DeMilio (2005)	Program participants (n = 84)	Other offenders (n = 84)	Diversion program	Recidivism	Citation/arrest	6 months	Chi square	TC offenders less likely to recidivate ($\chi^2 = 14.92$, $p < .001$)	
Norris et al. (2011)	Program participants (n = 635)	Historic (n = 186)	Diversion program	Recidivism	Referral	Varied: up to 4 years	Survival analysis	No significant differences between groups in survival curves	
Bright et al. (2013) (Baltimore City)	Program completers (n = 454)	Program non-participants (n = 332)	None	Recidivism	Referral	6 and 12 months	Descriptive	At 12 months: 16.6 % of completers had recidivated, compared to 43.7 % of non-completers and 42.7 % of referred non-participants	

Table 2 continued

Author (year)	Intervention sample	Control/comparison group	Comparison condition/program	Outcomes measured	Way recidivism measured	Measurement time frame	Analyses	Main effects	Sub-analyses
Bright et al. (2013) (Charles County)	Program completers (n = 403)	Program non-participants (n = 33)	None	Recidivism	Referral	6 and 12 months	Descriptive	At 12 months: 9.1 % of completers had recidivated, compared to 53.3 % of non-completers	
Bright et al. (2013) (Montgomery County)	Program completers (n = 710)	Program non-participants (n = 194)	None	Recidivism	Referral	6 and 12 months	Descriptive	At 12 months: 14.9 % of completers had recidivated, compared to 72.5 % of non-completers and 51.8 % of referred non-participants	
Jacobsen (2013)	Program participants (n = 31)	Other offenders (n = 174)	Diversion program	Recidivism	Conviction	Varied: up to 10 years	ANOVA	Mean recidivism of TC was 0.87 compared with 0.90 of comparison group 1 and 1.03 of comparison group 2 ($p = .840$)	
Laundra et al. (2013)	Program completers (n = 38)	Other offenders (n = 38)	Traditional justice processing	Self-reported behavior (intervention) vs. recidivism (comparison)	Referral	60 days	Descriptive	8 % of TC completers self-report being involved in a crime without being caught vs. 43 % of the comparison group charged with a second offense	
Norton et al. (2013)	Program participants (n = 223)	Program non-participants (n = 218)	None	Suspension; disciplinary infraction	–	Varied: ?	Descriptive	22 % of TC participants had another suspension compared to 29 % of those who were referred but did not attend	
Buchholz (2014)	Program completers (n = 87)	Other offenders (n = 286)	Traditional justice processing	Recidivism	Conviction	Varied: at least 1 year	Multivariable regression	Odds ratio of reoffending (TC vs. comparison) = 0.596 ($p = .367$)	

n.s. not significant

TC teen court

? Unable to be determined based on document(s) reviewed

Court). Despite these efforts, most still had limitations related to comparability, including offenders being able to select which (if any) type of diversion program they participated in, participation by different types of youth (e.g., race, gender, number and types of offenses), and different geographic areas or influences between the groups. The extent to which sufficient control variables were available and adjusted for in analyses varied greatly across studies.

While it was clear how youth were assigned in randomized studies, most quasi-experimental studies provided few details about how youth were selected to participate in the Teen Court program. Some studies mentioned that court officials screened participants based on characteristics of the youth or family and alleged offense, whereas others mentioned that youth or families were presented with options to participate in different processing mechanisms. The exact algorithms for assigning youth to programs were generally ill defined and when such protocols were mentioned, there appeared to be a large degree of subjectivity on the part of the Court, Probation, or Teen Court official. Rates of participation and program completion also varied across studies. For example, data in 2 counties examined by Bright et al. (2013) showed great divergence: in Baltimore County, 67 % of referred youth took part in a hearing and 57 % completed the program compared to Charles County, where 99 % took part in a hearing and 94 % completed the program. Although these studies were analyzed as two different Teen Court programs, many other studies pooled analysis of multiple court locations and did not disaggregate such statistics.

Studies varied in the types of programs or services received by the control/comparison group. Eight compared Teen Courts to traditional justice processing, seven compared Teen Courts to one or more alternative diversion programs, one compared Teen Courts to both a diversion program and traditional justice processing, and six compared Teen Courts to no intervention (youth who did not complete the Teen Court program or were not involved in the justice system) (Table 2). Most studies did not provide details about the types of interventions received in these alternative conditions. Some mentioned core components of alternative diversion programs (e.g., community service, restitution, linkages with services); such interventions primarily differed from Teen Courts in the inclusion of peers and the experience of a hearing. Less information was provided on traditional justice processing, which could range from a warning to an appearance before a juvenile court. This limitation was acknowledged by a few authors, including Stickle et al. (2008), who described the inability to fully describe the interventions received by youth in the comparison group as a limitation to their work.

Analyses from the 22 studies in the sample involved 7756 youth (mean = 353, standard deviation = 276). The

smallest study examined 76 youth while the largest examined 992 youth. Most study samples included a greater percentage of males than females and 16 of the 20 studies for which baseline population characteristics were provided included a majority of youth that were white. Only four studies included a majority of African American youth.

Outcome Measures and Analyses

Twenty studies included recidivism as an outcome measure. For 17 studies, recidivism was the only outcome assessed. Two of the 20 studies that examined recidivism also assessed attitudinal or behavioral measures, while 1 also measured residential placement. Recidivism was defined by 6 studies as any subsequent police contact (citation or arrest), by 11 studies as referral of a subsequent offense to court, and by 2 studies as conviction of a subsequent offense. Follow-up time periods for recidivism ranged from short amounts of time (60 days in one study, 5 weeks in another) to 18 months (2 studies) and 3 years (1 study). The most common time periods were 6 months (9 studies) and 1 year (5 studies). Seven studies used variable lengths of follow-up time (Table 2). Most studies mentioned obtaining data from juvenile files; only 1 study (Administrative Office of the Courts 1995) mentioned obtaining data on adult court involvement.

Six studies conducted only descriptive analyses for at least 1 of their primary outcomes. Eleven studies used unadjusted Chi square, *t* tests or analysis of variance as the test of statistical significance to measure the program effect for at least 1 of their primary outcomes. Four studies used multivariable linear or logistic regression models and 3 used survival analysis methods.

Direction and Size of Effects

Recidivism

Among the studies that examined recidivism, 4 found statistically significant results favoring Teen Courts, 1 found statistically significant results favoring processing of youth in the traditional juvenile justice system, 10 found null results, and 5 provided only descriptive statistics (did not report significance level) (Table 2).

The four statistically significant studies favoring Teen Courts were all quasi-experimental. All contained close to the average number of study participants, with three slightly below the average number. All had samples that contained predominantly males and whites, similar to most other studies. Two were completed as a part of the frequently cited Butts et al. (2002) research report which examined 4 Teen Court programs across the country. In the

Butts et al. (2002) report, juvenile offenders who participated in both the Alaska youth tribunal court and the Missouri youth judge court had significantly lower rates of recidivism (defined as 6-month delinquency referrals) compared to youth processed in the traditional juvenile justice system. The Alaska study used a historical comparison group whereas the Missouri study used a comparison group of youth processed by Family Courts in the same jurisdiction during the same time frame. Both identified comparison cases by randomly selecting youth matched on gender, race, age, and offense. Neither study described what (if any) interventions were provided to youth processed through the traditional juvenile justice system or provided data on process or fidelity assessments of the Teen Court model.

Hissong (1991) found a significantly lower rate of 18 month recidivism for Teen Court participants compared with those processed in the traditional juvenile justice system in a suburban community in Texas; however, how recidivism was measured was unclear. The court, described as an adult judge model, was one of the only studies in which the majority of referred offenses were noted as vehicle-related. The comparison group included youth processed in the same jurisdiction during the same time frame, matched based on offense, gender, age, race and zip code. Like Butts et al. (2002), Hissong (1991) provided little information on what interventions were offered in the comparison group and no data on process or fidelity assessments of the Teen Court model.

The study described by Forgays (Forgays 2008; Forgays and DeMilio 2005) found that Teen Court offenders were less likely to be charged with a crime 6 months after the Teen Court session when compared to youth processed through traditional court diversion. This court followed an adult judge model and conducted a fidelity assessment of its grounding in restorative justice principles. On average, youth participated in the program for 12 weeks. All youth in the sample were repeat offenders previously processed through a traditional court diversion program. The comparison group included first time offenders referred to traditional court diversion for a misdemeanor or gross misdemeanor, matched based on gender. The study lacked details on program attrition rates and although results showed variation in recidivism over the 3 year study period (31 % recidivism in year 1 in the comparison group vs. 80 % in year 3), reasons for this variation remain largely unexplained.

One study found statistically significant findings favoring processing in the traditional justice system. This quasi-experimental study by Povitsky (2005) examined one Teen Court in Maryland that had been in existence for 2 years. Little data was provided on the court model outside program participation criteria (age 11–18, first-time

misdemeanor offense, must admit guilt). Potential participants were given a choice of Teen Court or traditional justice system processing. In order to obtain 18 months of follow-up data, only youth ages 16.5 and younger were included in the study. Comparison data was drawn from offenders in a neighboring county who met age and offense severity criteria. The logistic regression model controlled for age, gender, race and offense; however, the potential for unmeasured differences between groups remained due to the non-randomized nature of the study design. The study was the largest included in this review.

Other Outcomes

Of the five studies that examined attitudinal or other behavioral outcomes, two showed at least some statistically significant effects, two showed null results, and one did not report significance (Table 2). Stickle et al.'s (2008) randomized study revealed significantly higher levels of self-reported delinquency among youth assigned to Teen Court compared to control group participants at post-test. The groups did not differ in self-reported drug use, social skills, beliefs, self-concept, rebelliousness, or neighborhood attachment. Logalbo and Callahan (2001) examined changes in attitudes over a 5-week period (before and after participating in an adult-judge Teen Court in Florida) by comparing Teen Court participants who completed the program with a convenience sample of local high school and middle school students. The study found that Teen Court participants had significantly greater increases in knowledge of laws and trial procedures, positive attitudes toward judges, and negative attitudes toward police over the course of the study period. The study failed to find differences in self-feelings.

Sub-analyses

Overall, analyses examining the differential impact of Teen Courts on sub-groups of youth or the effectiveness of different types/components of the Teen Court model were extremely limited. Hissong's (1991) study examined survival time by gender and found that male Teen Court participants survived (did not reactivate) significantly longer than males in the comparison group, but there were no differences for females. Wilson et al.'s (2009) analysis of data from Stickle et al.'s (2008) randomized trial also assessed differential impacts of the program on gender. Stickle et al.'s original trial found overall null results, with the exception of significantly higher self-reported delinquency among those assigned to Teen Court. When Wilson et al. (2009) considered gender, the effect sizes comparing treatment and control males were much larger than the effect sizes for females in drug use and delinquent

behavior; however, the interaction term between gender and program failed to reach statistical significance in the regression models.

With regard to programmatic components, two analyses by Butts et al. (2002) examined differences in Teen Court models (adult judge compared to peer jury) but did not find significant effects. Norris et al. (2011) examined the impact of the number of sanctions on recidivism and found that more sanctions were associated with increased recidivism; however, this effect was not statistically significant. Nochajski et al. (n.d.) examined the impact of a specific program component (a youth workshop focused on impulse management, values, goals, self-esteem, character education, and anger management); attendance was shown to have a significant negative effect on recidivism (odds ratio, workshop participants as the referent group = 23.58, $p < .001$).

Discussion

Despite the popularity of Teen Courts as a juvenile justice diversion model, evidence of their effectiveness is unclear. Current studies of Teen Courts, which vary widely in their program design and methodological quality, provide conflicting evidence of the impact of such programs on participating youth (Butts et al. 2002). This systematic review sought to address this gap by summarizing the state and quality of current empirical evidence. This systematic review found limited evidence to support the effectiveness of Teen Courts on short- or long-term outcomes for juvenile offenders. The majority of studies identified in this review, including two randomized controlled trials, failed to detect differences in recidivism rates over time when comparing Teen Courts to another form of juvenile processing. Of the 15 studies that assessed statistical significance of recidivism, 4 found statistically significant results favoring Teen Courts, 1 found statistically significant results favoring processing in the traditional juvenile justice system, and 10 found null results. The small number of studies that examined outcomes other than recidivism makes it difficult to draw inferences about the impact of Teen Courts on other attitudinal or behavioral outcomes. Given the large number of Teen Court programs in the U.S. and the number of agencies that have promulgated them as a promising model (Butts et al. 2002; Godwin 2001), it is surprising to find such limited evidence in support of their effect.

The present review revealed high levels of heterogeneity in all aspects of the current evidence base, including differences in intervention design, research design, and the extent of program monitoring and implementation. Furthermore, most studies lacked detailed descriptions of core intervention components, comparison conditions, and

outcome measures. This heterogeneity makes it difficult to compare interventions, quantitatively combine results, or identify best practices. This systematic review was not able to discern any patterns in court model/theory, participation criteria, study design, sample size, or outcome measures among the studies that did detect positive effects. Researchers and program implementers would be wise to heed the Office of Juvenile Justice and Delinquency Prevention's recent call for additional research to understand the effectiveness of Teen Court programs (Office of Juvenile Justice and Delinquency Prevention 2015). Based on this review, at least three key considerations with regard to a) study design and methods, b) reporting of the intervention and counterfactual conditions, and c) assessment of pathways and differential program effects should be taken into account in future studies (summarized in Table 3).

Study Design and Methods

Future studies should consider using stronger research designs. Of the 40 documents that were identified as including an empirical assessment of a Teen Court program on offender outcomes, 21 were included in this review because they described studies that used an experimental or quasi-experimental design. Of the 22 studies referenced in these documents, only two used randomization. Because the process of assigning youth to a juvenile justice processing mechanism is inherently subjective (i.e., based on an individual's appraisal of youth characteristics and needs), randomization of youth that meet clearly defined objective criteria would strengthen the potential for causal inference. In jurisdictions with a limited capacity to hear cases in a Teen Court, this may represent a particularly promising approach. If randomized designs are not possible, taking steps to describe the process of how youth entered into a Teen Court program (e.g., participation/screening criteria, whether youth are given an option to participate) and to select a suitable comparison group are paramount. Selecting a comparison group of youth who refused to participate, did not complete the program, or were not involved in the juvenile justice system presents serious threats to internal validity.

Consideration should also be given to how best to measure subsequent delinquency, the ultimate goal of juvenile diversion programs. The majority of studies in this review used an objective measure of recidivism; however, the exact meaning of the measure varied greatly, ranging from citations/arrests to referrals or convictions. Likewise, the timeframe of the measure varied greatly. The nature and timeframe in which recidivism is judged is likely to have a significant impact in how we understand the effectiveness of Teen Courts. Unfortunately, only 1 empirical study in the review assessed outcomes at more

Table 3 Recommendations for enhancing the quality of research and evaluation studies to examine the impact of Teen Court programs

Recommendation	Components
<i>Study design and methods</i>	
Use randomization or an adequately matched comparison group	<p>Randomize youth who meet program criteria to intervention or control condition. If randomization is not possible, take steps to match participants on key characteristics known to influence outcomes of interest</p> <p>Avoid selecting youth who do not attend a Teen Court hearing, do not complete the Teen Court program, or are not involved in the juvenile justice system as a comparison group</p> <p>Describe the process by which youth are assigned to the Teen Court program, including whether they are given the option to participate</p>
Examine different types of youth recidivism	<p>Consider multiple ways in which youth might have subsequent contact with the juvenile justice system</p> <p>Clearly define the type of recidivism being measured</p> <p>Examine recidivism at different time points</p> <p>Consider both subsequent contacts with the juvenile and adult justice systems in measures of recidivism</p>
Analyze impacts based on all youth assigned to intervention conditions	<p>Follow the intent to treat model by including all youth assigned to a program/condition in the analysis sample</p> <p>Describe program completion criteria and rates in both the experimental and control/comparison groups</p> <p>Take steps to assess and control for differential attrition rates</p>
<i>Reporting of the intervention and counterfactual conditions</i>	
Include key details to describe the Teen Court program	<p>Clearly describe core intervention components, including: the Teen Court model, the theoretical framework, program participation criteria (e.g., types of offenses, first offense, age, whether youth required to admit guilt), the length of time youth spend in the program, and the process and personnel responsible for monitoring sanctions</p>
Use and describe a comparison condition that represents a realistic practice alternative	<p>Use a comparison condition that represents the current state of practice in the jurisdiction</p> <p>Describe the core components of the comparison condition and the ways in which it differs from the Teen Court program under study</p>
<i>Assessment of pathways and differential program effects</i>	
Use process and short-term impact measures, in addition to long term measures	<p>Conduct process assessment to understand variation in implementation of the Teen Court program (e.g., across sites, over time) and fidelity to stated program logic framework or protocols</p> <p>Collect data on short-term knowledge, attitudinal, or behavioral measures and conduct path analysis in order to understand the way(s) in which Teen Court programs might influence long-term youth outcomes</p>
Conduct sub-analyses to identify differential impacts	<p>Examine the impacts of Teen Court program components or models on youth outcomes</p> <p>Explore potentially differential impacts for different types of program participants</p>

than a single time point and no studies examined differences in program effectiveness by comparing rates of arrests, referrals and/or convictions. Failure to consider the multiple ways in which youth may have subsequent contact with the juvenile justice system represents a significant gap in informing policy and practice (Council of State Governments Justice Center 2014).

Reporting of the Intervention and Counterfactual Conditions

Future studies should take additional steps to describe the core elements of both the Teen Court program under study and the comparison condition(s). As illustrated by the

present review and previous studies on Teen Courts (Garrison 2001; Nessel 2000), there is wide variation in the ways in which a program can be implemented, including the Teen Court model and structure of the hearing (e.g., level of youth involvement, goals of the hearing), types of juvenile offenders that are allowed to participate (e.g., age, type of offense, first-time or repeat offenders), and protocol for completing sentences (e.g., length of the intervention, overseeing party). All of these elements are likely to have implications for the effectiveness of Teen Courts in improving outcomes for participating youth. Such variations are likely to be magnified when studying multiple court sites or examining impacts over several years. A serious gap identified in the present review was a lack of

information on the dose (time period and intensity) of the intervention under study—only a third of the studies assessed provided any information on how long or how much contact youth had with the program after the initial hearing. Because the Teen Court hearing itself is likely to be a time-limited intervention (e.g., one 90 min session), understanding follow-up protocols may provide greater insight into potential effectiveness of the model. Without a clear understanding of what is implemented and explicit assessments of the differential success of these factors, it will be impossible to understand how best to structure Teen Court programs.

Process and fidelity evaluations, including assessments examining courtroom processes, will provide a cornerstone for building this new evidence. To date, the majority of observational work has been done from an ethnographic perspective (Ventura 2006; Bright et al. 2013) and examples of structured observation tools that assess the extent to which Teen Courts are aligned with stated program goals are limited (Greene and Weber 2008). Due to the potential for variation across court sites and over time, further development of such tools represents a significant need.

Tantamount to the critical appraisal of the Teen Court diversionary model is the need to accurately understand the types of programs or services received by youth in alternative conditions. Best practices for research specify the need to compare the intervention under study to alternatives routinely offered in practice (Berger et al. 2009). Whether Teen Courts should be compared to processing in the traditional justice system or other types of diversion programs likely depends on the protocols of a jurisdiction. Studies examining Teen Court program effectiveness should strive to clarify the content of these alternatives (e.g., goals, components, services) and describe the key ways in which they differ from the Teen Court model. The question of whether Teen Courts “work” necessitates the need to define “compared to what?”

Assessment of Pathways and Differential Program Effects

Additional work is needed to consider pathways of intervention effects. As described by Butts et al. (2002), Teen Courts can be grounded in as many as 6 theoretical underpinnings, including restorative justice, labeling, peer justice, and specific deterrence. Our ability to understand the relationship between these theories and offender outcomes is severely limited by the lack of rigorous studies that consider both short (e.g., attitudes, beliefs, knowledge) and long term (e.g., behaviors) outcomes. Additional studies that consider short term indicators and explore pathways are needed in order to support causal inference and the development of evidence-based practices.

A related gap is the need to conduct sub-group analyses to examine potentially differential impacts of Teen Courts on different types of participants. As suggested by Hisson’s (1991) and Wilson et al.’s (2009) work, Teen Courts may have different impacts on males and females, and may even be harmful for certain youth. Furthermore, significant racial/ethnic disparities are present in rates of juvenile arrest and incarceration (Annie E. Casey Foundation 2013). The lack of studies exploring whether Teen Courts are beneficial for non-white youth and youth with different risk profiles, backgrounds, and/or offense types (e.g., substance-related, violent) is one of the most striking research gaps uncovered by this systematic review.

Limitations

While this systematic review is one of the first to identify and synthesize data on the impact of Teen Court programs on juvenile offender outcomes, it has limitations. First, although efforts were made to identify and include all relevant sources, some may have been missed, especially those present only in the grey literature. Second, because of our desire to be expansive in describing the state the evidence base, we have included studies with weak quasi-experimental designs (e.g., use of non-program completers as a comparison group); caution should be taken when interpreting the results of such studies. Third, because of the wide variation among current studies, especially with regard to how recidivism was measured (e.g., different time points, examination of arrest vs. filing), it was not possible to conduct a formal meta-analysis. Fourth, this review was limited to the examination of the impact of Teen Courts on juvenile offender outcomes. It did not consider other potential benefits of Teen Courts, including impacts on youth who volunteer (as jurors, bailiffs, etc.) or surrounding communities.

Conclusion

This systematic review found limited evidence to support the potential impact of Teen Courts on reducing youth recidivism when compared to other types of interventions. This finding is tempered by a number of challenges in the current evidence base, including weak study designs, lack of description and assessment of intervention components, unclear and inconsistent outcome measures, and little examination of pathways or differential intervention effects. With increasing popularity and the resulting expansion of Teen Courts as a national model for juvenile justice diversion, the time is ripe for additional research to demonstrate the model’s range of possible impacts and to identify best practices. Additional studies which maximize

both internal and external validity, consider the pathways of intervention effects, and examine the potentially differential impacts of the program on an expanded range of youth outcomes are needed to help inform sound decision-making about this intervention at both the local and national levels.

Acknowledgments The authors thank Sebastian Thomas Lemire from the University of California, Los Angeles for his support and contributions to the project.

Authors' Contributions LNG conceived of the study, led its design, and coordinated and drafted the manuscript; TS led the review and contributed to the development of the manuscript; AD, MS, and TK contributed to the interpretation of the results of the review and the development of the manuscript. All authors read and approved the final manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Compliance with Ethical Standards

Conflicts of interest The authors report no conflicts of interest.

References

- Administrative Office of the Courts. (1995). *Report on the teen court programs in North Carolina*. Raleigh, North Carolina. Retrieved from Duke University School of Law Library.
- Administrative Office of the Courts. (2011). *Youth courts fact sheet*. San Francisco, California. Retrieved from http://www.courts.ca.gov/documents/Youth_Courts.pdf
- American Bar Association. (2007). *Youth cases for youth courts: desktop guide. A guide to typical offenses handled by youth courts*. Chicago: Authors.
- Annie E. Casey Foundation. (2013). *Youth incarceration in the United States*. Baltimore: Authors.
- Berger, M. L., Mamdani, M., Atkins, D., & Johnson, M. L. (2009). Good research practices for comparative effectiveness research: Defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force-Part I. *Value in Health*, 12, 1044–1052.
- Bright, C., Morris-Compton, D., Walter, J., Falls, B., & Young, D. (2013). *Multijurisdictional teen court evaluation: A comparative evaluation of three teen court models*. Maryland: Administrative Office of the Courts.
- Buchholz, M. M. (2014). *Effectiveness of restorative justice programs in the prevention of juvenile crime*. Master's thesis, North Dakota State University.
- Butts, J. A., Buck, J., & Coggeshall, M. B. (2002). *The impact of teen court on young offenders*. Washington, D.C.: Urban Institute. Retrieved from <http://www.urban.org/UploadedPDF/410457.pdf>
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company.
- Council of State Governments Justice Center. (2014). *Measuring and using juvenile recidivism data to inform policy, practice, and resource allocation*. New York: Council of State Governments Justice Center.
- Dugas, D. (2006). *Does the effectiveness of teen court depend upon gender?* Master's thesis, University of Maryland.
- Eden, J., Levitt, L., Berg, A., & Morton, S. (2011). *Finding what works in health care: Standards for systematic reviews*. Washington: D.C., National Academies Press.
- Evidence for Policy and Practice. (2007). *EPPI-centre methods for conducting systematic reviews*. London: University of London.
- Forgays, D. K. (2008). Three years of teen court offender outcomes. *Adolescence*, 43(171), 473–484.
- Forgays, D. K., & DeMilio, L. (2005). Is teen court effective for repeat offenders?: A test of the restorative justice approach. *International Journal of Offender Therapy and Comparative Criminology*, 49(1), 107–118.
- Garrison, A. H. (2001). An evaluation of a Delaware teen court. *Juvenile and Family Court Journal*, 52(3), 11–21.
- Godwin, T. M. (2001). *The role of restorative justice in teen courts: A preliminary look*. Washington: D.C., National Youth Court Center.
- Gough, D., Oliver, S., & Thomas, J. (2012). *An introduction to systematic reviews*. London: SAGE Publications.
- Greene, E., & Weber, K. (2008). Teen court juror's sentencing decisions. *Criminal Justice Review*, 33, 361–378.
- Hissong, R. (1991). Teen court—Is it an effective alternative to traditional sanctions? *Journal for Juvenile Justice and Detention Services*, 6(14), 14–23.
- Jacobsen, J. L. E. (2013). *The relationship between juvenile diversion programs and recidivism for juvenile offenders*. Master's thesis, University of Nevada.
- Laundra, K., Rodgers, K., & Zapp, H. (2013). Transforming teens: Measuring the effects of restorative justice principles in a teen court setting. *Juvenile and Family Court Journal*, 64(4), 21–34.
- Logalbo, A. F., & Callahan, C. M. (2001). An evaluation of teen court as a juvenile court diversion program. *Juvenile and Family Court Journal*, 52(2), 1–11.
- National Association of Youth Courts. (2015) Retrieved from <http://www.youthcourt.net/>
- National Research Council, Committee on Assessing Juvenile Justice Reform, Committee on Law and Justice, Division of Behavioral and Social Sciences and Education [National Research Council]. (2013). *Reforming juvenile justice: A developmental approach*. Washington DC: The National Academies Press.
- Nellis, A. (2011). *Addressing the collateral consequences of convictions for young offenders*. Champion Magazine, National Association of Criminal Defense Lawyers. Retrieved from <http://sentencingproject.org/doc/publications/Collateral%20Consequences%20NACDL%202011.pdf>
- Nessel, P. A. (2000). *Youth court: A national model, technical assistance bulletin No. 17*. Chicago: American Bar Association, Division for Public Education.
- Nochajski, T. E., Hayes, E., Kramer, L., Michaels, T., Hill, B., Schreck, L., Johnson, K., Sabino, A., Odell, M., Wiehe, A., Noonan, E., Dingman, M. (n.d.). *Hillside Children's Center: Livingston County youth court and community services evaluation*. New York: State University of New York, Buffalo. Retrieved from http://www.globalyouthjustice.org/uploads/Livingston_County_Youth_Court_Evaluation.pdf
- Norris, M., Twill, S., & Kim, C. (2011). Smells like teen spirit: Evaluating a Midwestern teen court. *Crime & Delinquency*, 57(2), 199–221.
- Norton, M. H., Gold, E., & Peralta, P. D. R. (2013). *Youth courts and their educational value: An examination of youth courts in Chester, Pennsylvania*. Pennsylvania: Research for Action. Retrieved from http://www.researchforaction.org/wp-content/uploads/publication-photos/1187/Norton_M_Youth_Courts_and_their_Educational_Value.pdf

- Office of Juvenile Justice and Delinquency Reforms. (2015). *Funding opportunities list*. Retrieved from <http://www.ojjdp.gov/funding/fundinglist.asp>
- Patrick, S., & Marsh, R. (2005). Juvenile diversion: Results of a 3-year experimental study. *Criminal Justice Policy Review*, 16(1), 59–73.
- Patrick, S., Marsh, R., Bundy, W., Mimura, S., & Perkins, T. (2004). Control group study of juvenile diversion programs: An experiment in juvenile diversion—the comparison of three methods and a control group. *The Social Science Journal*, 41, 129–135.
- Povitsky, W. (2005). *Teen court: Does it reduce recidivism?* Master's thesis, University of Maryland.
- Puzach, C., & Haas, A. (2014). Addressing the needs of juvenile offenders: A study of the Greene county Missouri teen court model. *International Journal of Social Science Studies*, 2(4), 113–122.
- Schwalbe, C. S., Gearing, R. E., MacKenzie, M. J., Brewer, K. B., & Ibrahim, R. (2012). A meta-analysis of experimental studies of diversion programs for juvenile offenders. *Clinical Psychology Review*, 32(1), 26–33.
- Seyfrit, C. L., Reichel, P. L., & Stutts, B. L. (1987). Peer juries as a juvenile justice diversion technique. *Youth & Society*, 18(3), 928–940.
- Stickle, W. P., Connell, N. M., Wilson, D. M., & Gottfredson, D. (2008). An experimental evaluation of teen courts. *Journal of Experimental Criminology*, 4(2), 137–163.
- Valentine, J., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods*, 13(2), 130–149.
- Ventura, H.E. (2006). *Restorative justice and youth courts: An examination of implementation intensity, recidivism, and accountability*. Dissertation, University of South Carolina.
- Weisz, V., Lott, R. C., & Thai, N. D. (2002). A teen court evaluation with a therapeutic jurisprudence perspective. *Behavioral Sciences and the Law*, 20(4), 381–392.
- Wilson, D. M., Gottfredson, D. C., & Stickle, W. P. (2009). Gender differences in effects of teen courts on delinquency: A theory-guided evaluation. *Journal of Criminal Justice*, 37(1), 21–27.