Is Restorative Justice Effective in the U.S.?

Evaluating Program Methods and Findings Using Meta-analysis

by

Kyle Ernest

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2019 by the
Graduate Supervisory Committee:

Kathleen Talbot, Chair
Scott Decker
Stacia Roosevelt

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

The presence of restorative justice (RJ) in the United States has grown steadily within the last five decades. The dynamics of RJ programs are meant to more holistically address the harms caused by crime in comparison to the traditional criminal justice system (CJS). Yet, evaluative research has provided inconsistent evidence of their effectiveness and the quality of empirical study has gone untested. The current study sought to fill the gaps within past research by examining how success has been measured, assessing the rigor of study methodology using the Maryland Scientific Methods Scale (SMS), and determining the impact of RJ programs on recidivism, victim satisfaction and restitution compliance using meta-analysis. A systematic search of past literature identified a sample of 121 studies whose dependent measures were coded, and methodological designs were rated using the SMS. Most studies failed to include community-based measures of success or measures which reflect the goals of RJ to undue harms and restore relationships. SMS scores were well distributed within the sample. Despite restricted sample sizes, meta-analyses used extracted data from 35 case-control, quasi-experimental and experimental studies to generate 43 unique treatment contrasts and 3 summary effects. Meta-analytic findings favored RJ treatment over CJS control groups across all dependent measures. Heterogeneity between subsequent arrest studies was scrutinized using subgroup analysis. The fewest subsequent arrests were associated with adult offenders, mandated participation, conferencing and hybrid programs, and the most rigorous methodologies. Findings support continued efforts to improve the methodological rigor of evaluations, targeted focus on specific program types and delivery characteristics. Future meta-analyses would benefit from the inclusion of non-

American RJ program evaluations to enlarge pooled sample populations and better detect moderating influences. Other suggestions for research design improvements include the use of more holistic and stakeholder-centric measures for success, use of continuous measures, and refined indicator variables for heterogeneity testing (e.g., crime type severity, characteristics of program fidelity). The author recommends continued use of these programs, specifically with adult offenders and incidents of serious crime toward a better understanding of the true impacts of RJ on stakeholders. More detailed results, study limitations and implications are discussed herein.

DEDICATION

To my parents, for your endless love, support, and unwavering faith in me.

To my family, for your patience during this process and your endless encouragement.

To my friends, for your positivity, comfort, and companionship throughout this journey.

To my husband, for uprooting your life, for your strength, your wisdom, and your boundless humor. You made me laugh during my darkest moments, kept me grounded when I felt untethered, and helped me remember that adversity breeds growth. Hoot you.

ACKNOWLEDGMENTS

Throughout the research and writing of this dissertation I have received a great deal of assistance and advisement. I would like to express my deep gratitude to my committee chairperson, Dr. Kathleen Talbot. As a mentor, you instilled in me the values of self-investment, self-improvement, and self-worth. You were always on my team and your counsel and expertise were invaluable to my professional growth and academic success.

I would also like to thank the other members of my committee, Dr. Scott Decker and Dr. Stacia Roosevelt. You taught me the importance of learning from my failures and becoming a utility researcher. Your thoughtful consideration of my work, advice on methodology, and recognition of the implications of my research were especially helpful.

Thank you to Shannon Stewart for your varied help throughout the program, Julia Mesler for your contributions to coding, and Natasha Pusch for your time in reviewing my work. I appreciate the broader CCJ faculty and department for creating an environment which fosters both collaborative and autonomous work.

In addition, I would like to acknowledge all of my parents, family, and friends without whom none of this would be possible. Thank you for teaching me to be present and maintain a work-life balance. Your genuine interest in and support of my work means the world to me.

Finally, I would like to acknowledge my husband for being my sounding board over the years. You know as much about this project as I do! Thank you for reminding me to not take life too seriously and encouraging me to pursue my goals wholeheartedly.

You have work so hard and sacrificed so much to help me achieve my aspirations, Andy. I am forever grateful to have you as my partner in life. I love you and I like you.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

LIST OF EQUATIONS

EQUATION Page

Introduction

Monetary Cost of Criminal Justice

The price paid for America justice is considerable. The collective expense of police, courts, and corrections has increased by more than 200 billion dollars since 1982, with a recorded 265 billion dollars in expenditures for the 2012 fiscal year (Kyckelhahn, 2011; Wagner, 2003; Wildra, 2017). These costs exceed federal spending on the departments of agriculture, education, energy, labor, and transportation (Office of Management Budget, 2018). Budgetary demands of the judicial branch continue to grow despite the implementation of cost containment initiatives in the early 2000s. In 2018, the Federal Judiciary requested more than 7 billion dollars from Congress to support court operations and services (United States Courts, 2014; 2018). With an inmate population *five times* larger than most other developed countries, the funds allocated for correctional costs are even greater (Pratt, 2018). Annual spending on the maintenance of correctional facilities combined with providing care for an aging offender population exceed 60 billion dollars (Anno, 1990; Henrichson & Delaney, 2012; Schmitt, Warner, & Gupta, 2010). Yet, the vast expense associated with supporting the criminal justice system is underscored by largely unsuccessful techniques for deterring criminality. Evidence of the impact of incarceration on crime reduction is not compelling (Goldkamp, 2003; Pratt, 2018; Reid, 2016; Shapiro, 1990). The moderate drop in overall crime in recent years cannot be attributed to the use of incarceration. Instead, research indicates that lowered crime rates are correlated with improved economic opportunities, modern policing strategies, and a decreased demand for illicit drugs such as crack cocaine (Brown, 2010; Western & Pettit, 2010). Though the fiscal burdens imposed by criminal justice are of great concern to policy-makers and funding agencies, they represent but one thread in a complex web of unintended harms. The treatment of offenders is also widely

criticized for being overly punitive and inequitable (Bennett, 2016; Everett & Wojtkiewicz, 2002; Western & Pettit, 2010).

Ethical Issues of Criminal Justice Practices

Increased and Disparate Use of Incarceration. Ethical concerns have been raised over the excessive use of incarceration and disparate imprisonment of racial, ethnic minorities. In the last four decades America has seen a more than 220% increase in its rate of incarceration. The number of people serving life sentences without the possibility for parole has more than quintupled since the mid-1980s (The Sentencing Project, 2018). Yet, the more frequent use of incarceration and elongation of prison sentences are not indicative of increased crime commission. Instead, scholars attribute the mass incarceration epidemic to the sentencing policies implemented by the "war on crime" and "war on drugs" movements of the late 20th century (Mauer, 2018; Reid, 2016). These campaigns were characterized by aggressive law enforcement and mandatory minimum sentencing which were employed discriminately (Engen & Steen, 2000; Petersillia & Greenwood, 1978). The reforms that followed, Mandated Sentencing Guidelines, then Advisory Sentencing Guidelines, were enacted to mitigate the subjectivity, often influenced by race, of sentencing-decision makers. However, the guidelines only served to further net-widening and racially skewed imprisonment (Everett & Wojtkiewicz, 2002; Kramer & Steffensmeir, 1993; McDonald & Carlson, 1993). Recent estimates of the current correctional population surpass 2.2 million individuals with approximately 4.5 million people under the supervision of probation or parole (Bonczar, Hughes, Wilson, Ditton, 2018; The Sentencing Project, 2018). Surprisingly, many convicted offenders are sentenced to prison for non-violent or victimless crime. Conservative estimates maintain that over a quarter of the correctional population is comprised of property offenders (27.5%) and nearly a third of inmates

(30.8%) are imprisoned for drug-related crime. Additionally, young people of color are overrepresented within the correctional system (Pratt & Godsey, 2003; Spohn, & Holleran, 2002). An overwhelming majority of all inmates are male (87.6%) or racial, ethnic minorities (56.7%) (Bonczar et al., 2018; Currie, 2013; Travis, 2014). Prior research suggests that the disproportional imprisonment of young, Black, males is symptomatic of discriminatory sentencing procedures (Cox & Rhodes, 1990; Nolan, 2003; Steffensmeier, Ulmer, & Kramer, 1998).

Influence of Extra-Legal Factors in Court Processes. The operations of the United States judicial system are greatly scrutinized for being biased. Disparate plea-bargaining and sentencing practices have negative implications for procedural justice. Perceptions of fairness for low-income or racial, ethnic minority defendants are consistently unfavorable. Research suggests this is due in part to their disproportionate likelihood of receiving more frequent and lengthier imprisonment sanctions (Casper, Tyler, & Fisher, 1988; Ogletree Jr, 1995; Tyler, 1984). Though excessive judicial discretion is credited with discriminatory sentencing, unchecked prosecutorial discretion has been associated with unethical plea-bargaining whereby prosecutors leverage their authority as decision-makers bolster their clearance rates (Burke, 2007; Devers, 2011; King, Soule, Steen, & Weidner, 2005). Low-income individuals and people of color may be most susceptible to this type of coercion. Research reveals that indigent and minority defendants are routinely pressured into guilty pleas to avoid trial (Kurlychek & Johnson, 2004; Steffensmeier & Demuth, 2000; Ulmer & Bradley, 2006). However, poor sentencing decisions are not always motivated by overt partiality or greed. While subjective sentencing practices may be an effect of implicit bias, practical limitations adversely influence court outcomes as well (Nolan, 2003; Rachlinski, Johnson, Wistrich, & Guthrie, 2008; Richardson, 2016).

The struggle to cope with time constraints is a real problem for judges. The time for judicial sentencing decisions has been significantly reduced because of the sheer number of cases being processed by criminal court (Bennett, 2016; Richardson, 2016). Although criminal sentencing may take hours or even days, often judges only have minutes to render a decision in order to keep-up with insurmountable caseloads. Equal consideration of the legal and situational factors involved in each case is simply not feasible in such short timeframes (Bennett, 2016; Spamann, & Klöhn, 2016). As a result, sentencing decisions often reflect less-informed and proportionally inappropriate sanctions based on extra-legal factors (Bennett, 2016; Everett & Wojtkiewicz, 2002; Rachlinski et al., 2008; Redlich, Bushway, & Norris, 2016). Legal characteristics like crime type, crime seriousness, and criminal history play the biggest role in sentencing decisions but extralegal factors, such as race, ethnicity, age, and employment status are known to unduly influence sanctioning process (Kim, Cano, Kim, & Spohn, 2016; Steffensmeier et al., 1998; Steffensmeier & Demuth, 2000; Spohn, 2000). Consideration of such extralegal factors to make sentencing decisions and over-reliance on incarceration as a means for crime control function to marginalize racial and ethnic minority individuals and their respective communities (Alexander, 2012; Clear, 2009; Pratt, 2018; Roberts, 2003).

Negative Impacts of Incarceration

The use of certain, swift, and severe punishment (i.e., deterrence) as a crime control strategy has not been fruitful. Incarceration has not only failed to effectively caution the general population against criminal engagement (i.e., general deterrents) but has also been lacking in its ability to produce desired behavioral changes in known offenders (i.e., specific deterrents) (Braman, 2007; Cerrato, 2014; Cullen, 1995; Cullen, Skovron, Scoot, & Burtin Jr., 1990; Roach, 2000). Despite moderate reductions to the overall crime rate in the last decade, high levels of

crime and recidivism persist. More than two-thirds of persons released from incarceration reoffend within 3 years (Alper, Durose, & Markman, 2018; Cullen, Jonson, & Nagin, 2001). This is especially true for drug crime (Harrison, 2001; Spohn & Holleran, 2002). These figures support the argument that avenues for offender rehabilitation, such as substance abuse counseling, are constricted within the correctional system (Bales, Van Slyke, & Blomberg, 2006; Phelps, 2011; Watson, Stimpson, & Hostick, 2004).

Obstacles for Offender Rehabilitation. Lacking rehabilitative services and inhumane conditions perpetuate criminality within and beyond prison walls (Cullen & Gilbert, 2012; Phelps, 2011; Simon, 2014; Singer, 1970). Instead of rehabilitating individuals and decreasing criminal opportunity, evidence suggests that the conditions of incarceration operate to increase future re-offending, exacerbate drug dependence, and hinder informal crime control within communities (Bales et al., 2006; Feeley & Simon 1992; Lynch & Sabol, 2004; Nagin, Cullen, & Jonson, 2009; Pratt & Cullen, 2000; Spohn, & Holleran, 2002). Correctional overcrowding increases inmate risk for mistreatment, physical victimization, and sexual abuse (Morin, 2013). Crime within the correctional system is ever-increasing, especially for drug or gang related crime (Gaes, Wallace, Gilman, Klein-Saffran, & Suppa, 2002). Drug offenses are the highest crime category in the correctional system accounting for about half (45.4%) of all crime committed in prison (Federal Bureau of Prisons, 2019). Violence between inmates and between correctional staff and inmates is also problematic (Patrick, 1998). These crime trends lead scholars to believe that prison is not safe or effective (Buchanan, 2007; Morin, 2013). Conditions upon re-entry present additional roadblocks to offender reform.

Obstacles for Offender Re-entry. The conditions of re-entry do not support successful reintegration. “Felon disenfranchisement”, a term used to describe the restriction of certain rights

for the formerly incarcerated population, is a serious problem in the United States. Disenfranchisement laws take many forms and vary by jurisdiction (Liles, 2006). The ineligibility of former inmates to vote is common, however other restrictions encroach on day-to-day life and create greater criminogenic risk for those recently released. Housing and employment are two of the greatest predictors for avoiding future recidivism (Parhar & Wormith, 2013). Yet, many people re-entering communities are subject to residency restrictions and employment regulations such as "check-the-box", effectively limiting their avenues for successful reintegration. These issues are especially pervasive for people convicted for sex or drug related crime (Duwe, 2009; Freeman, 2003; Walker, 2007). The formerly incarcerated population also tends to be socially isolated (Willis & Grace, 2009). Stigma produced by criminal labeling acts as a barrier to establishing and maintaining pro-social relationships that are critically important to avoiding recidivism (Chiricos, Barrick, Bales, & Bontrager, 2007; Cochran, 2014). Felon disenfranchisement laws and poor community reception of re-entering offenders contribute to high rates of re-offending and subsequent re-incarceration.

Negative Impacts on Communities. The "revolving door" of imprisonment and re-entry fosters concentrated disadvantage (Chamberlain & Hipp, 2015; Wehrman, 2010). Areas of concentrated disadvantage are characterized by criminogenic conditions (Kirk & Papachristos, 2017). By "locking-up" a population comprised of young, Black, males, communities are subjected to skewed gender ratios, lacking supervision of juveniles, and an overabundance of single-parent households (Alexander, 2012; Roberts, 2003). Furthermore, residential transiency, social disorder and isolation, and family disintegration make "legitimate means" scarce and perpetuate the cycle of crime (Clear, 2009; Kubrin, & Stewart, 2006; LaFree, 2018; Sampson & Groves, 1989). Research suggests that the negative impacts of community disenfranchisement

such as systemic poverty, poor health, and low collective efficacy are becoming generational (Alexander, 2012; Iguchi, Bell, Ramchand, & Fain, 2005; Roberts, 2003; Sloane & Choi, 2016; Wildeman & Western, 2010). Procedurally unjust treatment and the disproportionate incarceration of minorities has not only damaged community conditions but has also socialized citizens to question the credibility (i.e., legal cynicism) and utility of the criminal justice system (Gau, 2015). When police are seen as illegitimate and incompetent individuals are less likely to report victimizations (Bliz, 2007; Kirk & Papachristos, 2011; Yoon, 2015). Consideration of the the treatment of victims within the judicial system is also of concern.

Negative Impacts on Victims. Prior literature notes a growing dissatisfaction with the perceived ability of the western justice system to meet the needs of those affected by crime (Benesh, 2006; Cerrato, 2014). Research suggests that victim needs in the criminal justice system are not being sufficiently prioritized. Justice agents such as law enforcement and court actors work to process and win cases often before managing resources, including mediation, for the victims of crime (Bazemore & Mahoney, 1994; Umbreit, 1995). Additionally, victims have limited participation and influence over decisions to charge and the disposition of criminal cases. This is especially true for young, racial, ethnic minority victims of violent or sexual crime (Poe-Yamagata, 2009; Kaukinen, 2004; Spohn & Spears, 1996). Feelings of disempowerment, exclusion, and frustration are common among victims (Benesh, 2006; Cerrato, 2014; Kirchengast, 2016). Research also suggests that negative experiences for victims may be traumatic and hinder emotional recovery by increasing feelings of fear and anxiety (Kunst, Popelier, & Varekamp, 2015; Lens et al., 2015). This “re-traumatization”, which occurs as an effect of the justice process and not the crime itself, is especially common among victim survivors of sexual assault or rape (Seidman & Pokorak, 2011; Seidman & Vickers, 2005).

Alternatives to traditional court proceedings and incarceration have emerged to mitigate adverse outcomes for offenders, victims, and communities.

Alternatives to Incarceration

Criminal Justice Reforms. Empirical research thoroughly documents the failures of traditional criminal justice. A critical ideological shift from retributive justice toward more holistic justice frameworks is reflected in recent policy reform. The "Penal Harm" and "Nothing Works" movements have given way "Therapeutic Jurisprudence", a term used to encapsulate crime control strategies which de-emphasize formal contact with the criminal justice system and punishment (i.e., incarceration) and utilize alternatives to incarceration (e.g., diversion, community supervision) (Cullen, 2013; Martinson, 1974; Palmer, 1975). Therapeutic courts, also known as specialized courts, create opportunities for individualized treatment of offenders, provide rehabilitative services, and are often a vehicle for less formal sanctions like community supervision (Cayley, 1998). Juvenile crime, first time offenses, and drug related crime are most frequently diverted to specialized courts to avoid incarceration and mandate restitution and other services (e.g., psychiatric care, substance abuse counseling) (Warner & Kramer, 2009). The implementation of specialized courts and community corrections have yielded mixed results in terms of recidivism outcomes (Bonta, Rugge, Scott, Bourgon, & Yessine, 2008; Greenwood & Turner, 2011; Petitclerc, Gatti, Vitaro, & Tremblay, 2013). Though recidivism outcomes for drug court have been favorable, there is reason to believe that its impact is moderated by other offender characteristics (i.e., offender risk) (Mitchell, Wilson, Eggers, & MacKenzie, 2012; Sloan, Smykla, & Rush, 2004; Spohn, Piper, Martin, & Frenzel, 2001). Specialized courts have also resulted in unintentional net-widening whereby more individuals are diverted to specialized courts who otherwise would have had limited contact with criminal justice, effectively increasing

the number people under supervision of the system (Gross, 2010; Hoffman, 2000, 2017). "Restorative justice" has also been adopted as an alternative to traditional court processes and incarceration.

Restorative Justice. Restorative justice aims to undo the harms imposed by interpersonal conflict (i.e., crime), rehabilitate offenders, and repair damaged relationships between crime victims, offenders, and communities (i.e., stakeholders). In the United States restorative justice is most commonly used with juvenile delinquents and less serious or first-time offenders. Restitution was the first recognized use of restorative justice in the United States during the 1970s (Barnett, 1977). It is used to divert juveniles from formal justice processes while creating practical obligations to "make-right" on their criminal wrong-doing (Eglash, 1958; Schneider, 1985). Restitution takes different forms such as monetary fees, work agreements, and community service. Other forms of restorative justice are characterized by shared dialogue (i.e., mediation, conferencing, circles) (Sullivan & Tifft, 2008; Zehr, 2015). Unlike incarceration which uses incapacitation to hold offenders accountable, restorative justice fosters offender acknowledgement of wrong-doing in more personal, action-based ways, helping victims and communities of crime obtain justice which may be more meaningful than retributive punishment (e.g., story-telling, apologies) (Johnstone, 2013; Bazemore, 1998; Walker, 2006). Restorative justice is also lauded for its inclusion of victims and communities as key decision-makers.

Where criminal justice is a superficial solution to crime, imposing sanctions based on an offense, restorative justice directly addresses the root causes of crime and creates individualized approaches for undoing specific harms. Despite the wealth of literature detailing the victories and failures of traditional criminal justice, by comparison lesser scholarly attention has been dedicated to assessing the impacts of restorative justice programs. Empirical research utilizing

data from the United States is especially scarce (Beale, 2003; Kurki, 2000). The results of existent evaluative literature are limited by issues of inconsistent definitional and conceptual components, lacking agreement of its goals, use of varying and unreliable dependent measures, and methodological flaws (Bazemore & Day, 1996; Bergseth, & Bouffard, 2007; McCold & Wachtel, 2002, 2003; Schiff, 1998). Even rigorous evaluations suffer from these deficits, which makes drawing conclusions about the effectiveness of restorative justice an impossible task (Bonta, Jesseman, Rugge, & Cormier, 2006; Poulson, 2003). These factors make clear the need for greater empirical attention in this area to decide how best to proceed in our future efforts to solve the problem of crime within the United States.

Purpose of the Current Study

As the use of restorative justice becomes more common in the United States the importance of sound evaluation methodology is more necessary than ever before. The current study addresses the weaknesses of past research in several ways. This study utilizes empirical research on restorative justice programs including restitution, mediation, conferencing, and circles specific to the United States. The systematic exclusion of foreign data improves on prior meta-analyses which have relied wholly on international samples (Bradshaw & Roseborough, 2005; Latimer, Dowden, & Muise, 2001, 2005), or have focused exclusively on one treatment modality of type of offender (Nugent, Williams, & Umbreit, 2003, 2004). Furthermore, only the highest quality case-control, quasi-experimental and experimental research is included in meta-analytics. Scholars have raised concerns over the issues of self-selection bias and poor randomization within impact evaluations but have otherwise failed to control for these issues within meta-analyses (Bergseth & Bouffard, 2013; Bouffard, Cooper, & Bergseth, 2017). The current investigation employs an objective global ranking scale in its consideration of sample

selection based on methodological quality. Using this unique sampling strategy, the current study compared dependent variable measures, quality of methodology, and estimated effect sizes for recidivism (i.e., subsequent arrest), victim satisfaction, and restitution compliance outcomes. Heterogeneity within the sample is also examined post-hoc by using subgroup analysis that explored variance in treatment effects due to different aspects of program delivery and program type.

The results of this investigation help to clarify the effectiveness of restorative justice in the United States as success has been measured in prior research. Results have implications for justifying the use restorative justice in its different forms to meet the needs of stakeholders and decrease offender recidivism. The following chapter provides an overview of restorative justice, including its origins, theoretical foundations, differing forms, and use in America. The literature review chapter also presents a scholarly debate about what constitutes restorative justice and success as well as synthesis of extant meta-analyses and their shortcomings. Finally, chapter 2 details research questions and hypotheses. The methods chapter sets the parameters of the current study including the sample selection process, systematic data collection, coding of methodological strength and dependent variable measures, and meta-analytic and subgroup analysis procedure. The descriptive findings pertaining to methodological rigor and outcome measure coding for 121 evaluations are outlined in chapter four in text and table format. The results chapter also provides the meta-analytic and heterogeneity test findings for 35 case control, quasi-experiments and experiments for subsequent arrest, victim satisfaction, and restitution compliance outcomes. A total of 43 independent effect sizes and combined treatment effects were estimate for each dependent variable. Subgroup analysis and publication bias findings are also presented in the results chapter. Chapter five discusses the limitations for the

current study as well as implications for restorative justice practices in the United States and offers suggestions for future evaluative research.

Literature Review

Defining Restorative Justice

Many monikers have been assigned to the movement toward a more holistic, therapeutic response to crime in America. *Community justice*, *peacemaking justice*, and *transformative justice* are often used interchangeably with the term restorative justice in practice and research (Harris, 2004; Kurki, 2000; McCold, 2004; Wozniak, Braswell, & Vogel, 2008; Zehr, 2011). However, the term *restorative justice* best encapsulates a distinct framework which conceptualizes the problem of crime, its consequences, and solutions, in a fundamentally different way than the traditional criminal justice system (McCold, 2000; Mika & Zehr, 2003; Zehr, 2015). Instead of responding to crime as the breaking of a codified law, restorative justice is concerned with addressing "interpersonal violations" or "interpersonal conflicts" such as physical altercations, family disputes, or damaged property (Marshall, Fairhead, Kingsley, & Murphy, 1985; Sullivan & Tifft, 2008; Zehr, 2015). Instead of characterizing individuals as criminals to be held accountable by formal punishment, restorative justice tasks an offender, or "wrongdoer", with the responsibility of "making-right" the specific harms they have imposed (Brunk, 2001; Johnstone, 2013; Zerh, 2015). Whatever the conditions of "earned redemption", these practices emphasize an individualized approach to rehabilitation and facilitate the improvement of relationships between offenders, victims, and communities (Rodriguez, 2007).

Principles of Restorative Justice Practice. At their core, restorative justice practices are characterized by their inclusion of those impacted by crime in a process that helps undo harms and repair relationships (Braithwaite, 2004). The broad aims of restorative justice are victim recovery, offender rehabilitation, and the restoration of relationships (i.e., community connectiveness) (Weitekamp & Kerner, 2012; Weitekamp & Parmentier, 2016). There are three main principles of restorative justice practice (Johnstone, 2013; Zehr, 2015). First, *harms and*

needs, not only of victims but of all *stakeholders*. Regardless of the specific restorative program employed, the participants, those who have a 'stake' in addressing an interpersonal conflict or crime, are constant. The three stakeholders, also known as the three "pillars", whose participation is essential to most processes of restorative justice are the victim, offender, and community (Zehr, 2015). Although restorative restitution and certain forms of mediation may only involve offender and victim, all other dialogue-based programs necessitate the inclusion of community as represented by social supports (e.g., family, peers), citizens who are intimately involved (i.e., co-victims) or impartial citizens (i.e., resources for offender), or less frequently, legal representatives (e.g., lawyers, victim advocates, law enforcement, judges) (Sullivan & Tifft, 2008; Willis & Grace, 2009). Prioritizing their collective involvement helps to identify the unique and varied needs of each stakeholder in a way that holds them in equal regard (Zehr, 2015).

The second principle, *obligations*, refers to the responsibility of offenders and communities to work toward repairing harm. Restorative approaches establish 'doable' agreements to repair the harm imposed on victims, as well as communities, including conditions that are not related to restitution alone (i.e., earned redemption) (Bazemore, 1998; Johnstone, 2013; Sullivan & Tifft, 2008; Zehr, 2015). Examples of earned redemption include written apologies and participation in mentorship or problem management programs (e.g., substance abuse counseling, anger management) (Bazemore, 1998). The coordination of communities and offenders through the mechanism of shared obligations is complicated. Because there is some onus on communities for individuals' wrongdoing (i.e., a characteristic of a broken relationships), communities may play the role of indirect-victim as well as co-conspirator (Mika & Zehr, 2003; Zehr, 2015). Both roles having direct implications for achieving earned

redemption through meeting obligations (Bazemore, 1998). The third, and final principle of restorative justice practice is the active *engagement* of stakeholders in the resolution process. Stakeholders are engaged when they are encouraged to participate in the expression of harms (e.g., story-telling, questioning) as well as contribute to a collective resolution or obligation agreement (Zehr, 2015).

Definitional Components. For decades scholars have struggled to find common ground in terms of what constitutes restorative justice (Luna, 2003; McCold & Wachtel, 2002, 2003; Walker, 2013). There are broad definitions which are conceptually grounded. These are said to satisfy 'theoretical definition components' and draw heavily from the three principles of inclusion, obligation, and engagement (Briathwaite, 2004; Sullivan & Tifft, 2008). More narrowly focused definitions help to categorize specific programs by identifying their unique restorative elements, such as the varying participants (i.e., stakeholders), express goals, and protocols (Daly, 2017; Marshall, 1996). The latter type of definition satisfies 'process definition components' (McCold, 2000; Sullivan & Tifft, 2008). Reconciling these nuanced distinctions is challenging. However, one question bridges the abstract divide between these definitions: *does justice work to repair or restore the harms and damaged relationships caused by interpersonal conflicts or criminal offenses?* The current study uses this question to ground its operational definition of restorative justice which attempts to satisfy both theoretical and process definitional components. Programs are considered to meet the working definitions if: 1) programs self-identify as restorative justice, mediation, conferencing, circles, or restitution, and/or, 2) programs reflect efforts to bring together two or more stakeholders of a crime (i.e., victim, offender, community member(s)) to talk about an offense, harms caused by crime, or to create agreed upon obligations, in varying forms, for the purposes of offender redemption or victim recovery. By

using a definition of restorative justice, which is inclusive of diverse program criteria, the sample for the current study is representative of the full scope of evaluative restorative justice research while maintaining the basic principles of practice. This definition also allows for the inclusion of programs which conceptually represent restorative justice (i.e., encompass restorative principles) but may not identify as such (Walgrave & Bazemore, 1999).

Theoretical Foundations of Restorative Justice

Despite arguments that this framework is not guided by a common ideology, the core components of restorative justice are consistent with several criminological theories (Bazemore, 1996; McCold, 2000). Both social disorganization theory and social bonds theory correspond with the primary goals of restorative justice to repair relationships and restore harms (Braithwaite, 2004; Wheeldon, 2009). However, the two theories most closely associated with restorative justice are labeling theory and shaming theory (i.e., reintegrative shaming theory) (Becker, 1960; Braithwaite, 1989; Harris & Burton, 1998; Hay, 2001; Luna, 2003; Lynch & Sabol, 2004; McCold & Wachtel, 2002; Makkai & Braithwaite, 1994; Sampson & Grove, 1989; Shaw & McKay, 1942). Labeling theory originates from sociology, but it has been appropriated and applied to research in the fields of criminology, psychology and medical science (Link, Cullen, Struening, Shrout, & Dohrenwend, 1989; Moore, 2001). The theory stipulates that the assignment of negative labels, such “deviant”, “delinquent”, or “criminal”, influence self-identity and, in-turn, conformity to undesirable behavior (Becker, 1960).

Labeling Theory. People conform to their criminal labels (Becker, 1960). Through the process of symbolic interactionalism, criminal stereotyping and labeling by authority figures effectively warps self-identity, especially for juveniles in formal contact with the criminal justice system (Bernburg, 2009; Klein, 1986; Schur, 1971). Negative labels have long-lasting

repercussions. “Criminals” and “Ex-Cons” are stigmatized and met with suspicion and distrust by their communities upon re-entry (Winnick & Bodkin, 2008). Research suggests that labels are discriminately applied to preserve the power structure of society by oppressing minorities (Alexander, 2012). Restorative justice disrupts the process of negative identify formation by using the term ‘wrongdoers’ which is a more humanizing alternative to “criminal”. Using this label results in better self-esteem as well as less stigma and social isolation, all of which give individuals a better chance at successful rehabilitation and reintegrating (Braithwaite, 1999; Farrington & Murray, 2013; Makkai & Braithwaite, 1994).

Shaming Theory. Reintegrative shaming is the mechanism by which restorative justice reduces re-offending or undesirable behavior (Briathwaite, 1989). Shaming or *reintegrative shaming* theory barrows ideas from subcultural, control, opportunity and social learning theories of crime, but is mostly derivative of labeling theory (Bazemore & Schiff, 2015; Braithwaite, 2004; Wilson et al., 2017). Shaming acts as a positive socialization that weakens ties to criminal institutions and relationships by confronting wrongdoers with the consequences of their actions for crime victims (Beck, 1997; Davis, 2009; Zehr, 1990). The process of shaming, and the swiftness of shaming, is integral to reducing criminal behavior (e.g., shame, remorse) (Braithwaite, 1989; Bazemore & Umbreit, 1995). Unlike traditional criminal labeling which has long-term repercussions, within the context of restorative justice, the term wrongdoer is temporary. Once obligations are fulfilled, the wrongdoer is *redeemed* and can successfully re-integrate into a community (Braithwaite, 1989; Braithwaite & Roche, 2001). The rationale behind this theory is that shaming, when limited, fosters successful offender rehabilitation and re-entry by cultivating feelings of empathy and remorse (Braithwaite, 1998, 1999; Braithwaite & Mugford, 1994; Harris, Walgrave, & Braithwaite, 2004; Jackson, 2009; Pranis, 1997; Reiss &

Tonry, 1986). These theories illustrate how the issues created by punitive sanctions outweigh their potential benefits. Restorative justice programs are used to shame wrongdoers in ways that are not debasing or long-lasting (Braithwaite, 1989; Winnick, & Bodkin, 2008).

Restorative Justice Program Types and Applications

There is no one-size-fits-all restorative justice program. Rather, different forms of restorative justice are employed based on situational factors (i.e., stakeholder participation) and practical limitations (Johnstone, 2013; Presser & Van Voorhis, 2002; Schneider, 1985). Though there are many practices that fit within the domain of restorative justice, the primary program types are *restitution*, *mediation*, *circles*, and *conferences* (Sullivan & Tifft, 2008; Zerh, 2015). Subcategories of programs within these domains were created overtime to suit different crime types, stakeholder needs, and pragmatic constraints. For instance, the broad restorative circles model includes more specific programs such as sentencing circles and circles of support and accountability (Duwe, 2013; Wilson, McWhinnie, Picheca, Prinzo, & Cortoni, 2007; Sullivan & Tifft, 2008). The degree to which programs are *purest* or most closely reflect restorative justice definitional and operational components varies (McCold & Watchtel, 2003; Mika & Zehr, 2003). Moreover, programs may claim to be restorative, but fail in practice to meet 'process' definitional components (Bazemore, 1996; Belgrave, 1995; Dooley, 1995; Marshell, 1996; Zehr, 2015). Understanding what makes a program 'restorative', as well as the breadth of program types, is fundamental to assessing their impact.

Restitution. Restitution is the most well-established, and oldest, form of restorative justice in the United States (Eglash, 1958; Barnett, 1977). Restorative restitution, also known as balanced restorative justice, obligates offenders to exert efforts to make right their wrongs and to be rehabilitated in the eyes of the community through action (Fogel, Galaway, & Hudson, 1972;

Hudson & Galaway, 1975). Unlike dialogue-based programs, restitution does not require stakeholders to meet in-person, however such meetings are encouraged to establish restitution agreements (Johnstone, 2013). Restitution takes different forms including financial victim compensation, work agreements, community services, or personal restitution (e.g., fixing damaged property, apology letters) (Bibas & Bierschbach, 2004; Bazemore & Maloney, 1994; Galaway & Hudson, 1990; Sullivan & Tifft, 2008). Dissimilar to other forms of restorative justice, restitution may be mandated by formal justice agents, especially orders of restitution as a method of diversion. Pure forms of restorative justice programs endorse voluntary participation of all stakeholders, therefore mandated restitution is not ideal (Bibas & Bierschbach, 2004; Chappell, 2018; Zehr, 2015). However, it is commonly used in juvenile court as an alternative to incarceration and is an option when shared dialogue is inappropriate or not feasible.

Mediation. As the foundational model, mediation is considered the purest form of restorative justice. While restitution is the oldest form of restorative justice to go "mainstream" in America, deeply rooted in tribal origins such as the Māori Tribe of New Zealand, *community mediation* was the first statement driven form of restorative justice in the Unites States (Sullivan & Tifft, 2008). In general, there are four components that make-up the mediation process: (1) facilitated discussion, usually by a trained volunteer, between the wrong-doer and the victim, (2) shared dialogue about the harms created by crime, (3) collective creation of an agreement or contract, tailored to meet the specific, unique needs of all stakeholders, and (4) creation of a follow-up or inquiry about compliance (Johnstone, 2003; Umbreit, 1997, 2002). Deeply rooted in the its tribal origins,. Unguided by the theoretical framework that came later, community mediation has been used to manage various kinds of conflict, like non-criminal disputes, such as residential (i.e., landlord and tenant disputes), merchants (i.e., clients and businesses), and

institutional (i.e., educational institutions and pupils) (Sullivan & Tifft, 2008; Zehr, 2015). Diversion of criminal cases, or mediation in addition to formal sanctions, to *community-based dispute resolution* began in New York in the late-1970s. The goal of these programs was not to avoid recidivism, but to fulfill more 'justice' for victims (Barnett, 1977). Other states soon followed suit and by the mid-1980s more than 400 conflict resolution centers were using community mediation to reconcile informal and criminal cases (McCold, 2008).

Another form of mediation, *victim offender reconciliation* programs, emerged in Canada during this time. These programs were largely popular within the Mennonite community and spread from Ontario, Canada into the United States (e.g., Indiana, California). The main distinguishing feature of this program type was the expectation of forgiveness on behalf of victims (Zehr, 1990). Unlike legalistic forms of mediation, which tend to be statement-driven and arbitrative in nature, victim offender reconciliation is more dialogue-driven and focuses on facilitating communication between victims and offenders without prioritizing punishment. The parties must reach the resolution together, after they both have shared their personal feelings about the crime. However, the religious orientation and perceived pressure placed on victims is often criticized by scholars (Van Ness & Heetderks, 1997). In efforts to address these concerns a new form of mediation was developed. *Victim offender mediation*, similar to other mediation forms, includes face-to-face contact between victims and offenders to address the harms imposed by crimes. Victim offender mediation was considered less faith-based and diverged from earlier mediation types in two ways: (1) incorporating standardized training of facilitators or employing social workers as facilitators and, (2) expressing offender remorse or victim forgiveness. These were not express goals of the process, but rather potential byproducts (McCold, 2008; Umbreit, 1997).

Circles. Circles have strong aboriginal roots in New Zealand (i.e., Mauri Tribe), Canada (i.e., First Nation), and the United States (i.e., Navajo Nation) (McCold, 2008; Sullivan & Tifft, 2008; Zehr, 2015). Reflective of tribal culture values, restorative circles (i.e., *healing circles*) are underpinned by the concept of autonomy, not imposing sanctions. Conversely, *sentencing circles* emphasize a community model of justice, where elders or a sentencing authority make recommendations for resolution which incorporates the efforts or continued contact between multiple stakeholders (Byrd, 2008; Van Ness & Heetderks, 1997). These protocols rarely involve formal agents of justice. While typical function of circles is to bring community together, including victim and offender, to resolve conflict, however this model has been adapted to serve different purposes such as support groups (e.g., substance abuse, victim support, offender support) (Sullivan & Tifft, 2008).

Several restorative circle models exist (Bazemore & Umbreit, 2001). Peacemaking circles, healing circles, and sentencing circles are used to repair harms in various inter-personal conflicts and criminal offenses. As an alternative to victim compensation, peacemaking circles, a traditional form of Navajo conflict resolution, focuses on restoring damaged relationships through dialogue (Yazzie, 1994). Peacemaking courts for Navajo people emerged in late 1990s. Consequently, the Navajo nation reserve has since become the biggest restorative justice jurisdiction in America (Byrd, 2008; Yazzie, 1994). The Hollow Water tribe, specifically the Ojibwa women, developed healing circles as a societal response to violent and sexual crime that, along with alcoholism, was prevalent in the 1980s (McCold, 1999; Bushie, 1997; Sullivan & Tifft, 2008). Sentencing circles bridge communities and formal justice system. Basic circle protocols, respectful dialogue between victims, offenders, and community members, are inclusive of justice agents such as law enforcement, attorneys, and judges (Pranis, 1997). One

subtype which follows this trend is sentencing circles. Those used in Minnesota and Vermont are typically overseen but impartial community representatives from the criminal justice system who are not involved in the specific crime and are held in a courtroom. This unique form of restorative circles blends traditional deterrent sanctions with more rehabilitative sanctions (i.e., community service) (Umbreit, Coates, & Vos, 2007). Lastly, *circles of support and accountability* use the peacemaking model, but provide offender with a clear mentor to ease the process of reintegration. These mentors meet with offenders on a regular basis and act as an informal, social support (Johnstone, 2013; Wilson et al., 2007).

Conferencing. Similar to circles, restorative conferencing involves direct contact between stakeholders, and potentially stakeholders' families, friends, or community peers for support, to negotiate unique offender obligations (Sullivan & Tifft, 2008; Umbreit, Coates, & Vos, 2002; Zehr, 2015). Contemporary restorative conferencing derived from the original conferencing practices used by the Mauori people of New Zealand and Australia. Restorative conferencing is predominantly used in the United States as a diversion resource for minor, juvenile, or first-time offenses (Choi et al., 2012; McCold, 2001), however some states also use this practice in cases of child-welfare, or with adult and serious offenders (Hudson, 2002). Types of restorative conferencing differ in terms of participants, facilitators, and who approves the agreement or contract (Sullivan & Tifft, 2008; Zehr & Umbreit, 1982). As implied by the name, *family group conferencing*, or *family unit meetings*, include the participation of families and victims as a key feature of the restorative process. These meetings are semi-structured. In juvenile justice cases the family members of offenders and victims contribute their own opinions as well as act as support. This model varies from the *family group decision-making conference*, which is more appropriate to use in cases concerning child-welfare (Zehr, 2015). Family

violence and child safety concerns are discussed between family members, official authorities and a facilitator (i.e., restorative social worker) to reach a conclusion about care and responsibility of the child.

In several countries these conferences have become standardized practice in all cases of domestic violence (i.e., intimate partner violence) and with child protection agencies. *Police conferencing*, based on the New South Wales and Australian models, is considered a form of community, and problem-oriented, policing (Sullivan & Tifft, 2008). Police are trained as facilitators and carryout conferencing guided by scripts with prompts for each stakeholder. Finally, *community conferences* are employed in both school and organizational environments to achieve conflict resolution. This form of conferencing also uses a prescribed script based on restorative ideals surrounding reintegrative shaming (McCold, 2008). There exists a wide variety of program types. Scholars have questioned the degree to which programs are truly restorative based on definitional and conceptual components (McCold & Wachtel, 2003). Further research is needed to qualify the impact of different restorative program types.

Hybrid Programs. *Hybrid* restorative justice programs, also known as *blended* or *balanced* restorative models, is a "catch-all" category of programs which offer multiple treatment modalities or integrate restorative elements into existing non-restorative practices (e.g., psycho-educational programs, specialized court processes) (Bazemore, Pranis, & Umbreit, 1997; Bazemore & Umbreit, 1994). These programs are typically used in juvenile diversion schemes as a way to ensure juvenile participation regardless of other willing participants (i.e., use non-dialogue based programs) or in school settings where different forms of restorative justice (e.g., peer mediation, family group conferencing, peacemaking circles) can be used to address problems (e.g., bullying, stealing, vandalism) (Bloch, 2010; Gilbert, Schiff, & Cunliffe, 2013).

However, the label of *balanced restorative justice* is most commonly used within criminal justice settings, such as specialized courts (Bazemore et al., 1997; Bazemore & Umbreit, 1994; Fronius, Persson, Guckenburg, Hurley, & Petrosino, 2016).

Prevalence of Restorative Justice

Though 'restorative justice' was officially coined in the 1950s its origins are rooted in ancient indigenous cultures (Eglash, 1958; Johnstone, 2013; Van Ness, 1989; Van Ness, Carlson, Crawford & Strong, 1989). Peacemaking and conflict resolution evolved from practices of the Navajo Nation (Brunk, 2001; Sullivan & Tifft, 2008; Van Ness, 1993; Weitekamp & Kerner, 2012). The Mauori tribe of New Zealand and aboriginal societies of Canada commonly used this collaborative, communal dialogue in response to wrongdoings (Galaway & Hudson, 1990; Zehr, 2015). As a common alternative to traditional crime control approaches, more than a dozen countries utilize restorative justice processes (Hughes & Mossman, 2001; Johnstone & Van Ness, 2004; Sullivan & Tifft, 2008). Instead of strictly imposing retributive sanctions or solely diversionary or rehabilitation resources, policymakers from these countries view restorative justice as a viable alternative that may encompass elements from both justice models (Johnstone, 2013; Weitekamp & Parmentier, 2016). Australia, Africa, England, Canada, and the United States, among others, have all adopted some form of restorative justice (Galaway & Hudson, 1990; Hughes & Mossman, 2001).

The utility of restorative justice in the United States is underpinned by a notable shift toward a more rehabilitative, individualist approach to crime control (Freeley & Simon, 1992; Listwan, Jonson, Cullen, & Latessa, 2008; Marshall et al., 1985; Umbreit, 1985; Zehr, 1985). Restorative practices are heavily used within the juvenile justice system around the world. For instance, New Zealand has reformed their juvenile justice practices to be exclusively restorative,

accounting for all delinquency except in cases of homicide or rape (Zehr, 2015). Similarly, restorative justice in America is most commonly used with juvenile or first-time offenses to avoid incarceration (e.g., community service, restitution) (Roach, 2000; Umbreit, 1994). Though referrals by court, probation, and parole into diversion programs is not limited to juveniles, the use of restorative justice programs with adult offenders remains minimal in the United States (Bazemore & Umbreit, 1995; Choi**,** Bazemore, & Gilbert, 2012; Consedine, 1995; Immarigeon, 1999; Pranis, 1997; Umbreit & Greenwood, 1997; Van Ness, 1999). Restitution was the first form of restorative justice to take root in the United States in the 1970s. Since its adoption, other forms of restorative justice, such as victim-offender mediation, peacemaking circles, circles of support and accountability, victim impact panels, and restorative conferencing have emerged (Braithwaite, 1999).

Restorative justice conferencing and diversions have been integrated into existing juvenile justice systems within numerous states, including Florida, Idaho, Montanta, New Mexico, Colorado and more (Bazemore, 1997; Bazemore, 1998). Vermont has also widely adopted restorative justice sentencing boards for most non-violent crime (Belgrave, 1995; Karp, 2001), including felony crimes (Dooley, 1995). In these cases, sanctions are collectively decided through community panels, but practices such as mediation may be used after sanctions are ordered by judges, before or after incarceration (Belgrave, 1995). The adoption of these programs corresponds to movements for political and cultural change as they relate to decreasing retributive, punitive punishments in favor of more rehabilitative sanctions (Cullen, 2005; Cullen, & Gendreau, 2001). These examples illustrate increased interest and support of restorative justice practices in the United States (Bazemore, 1998; Bazemore & Umbreit, 1995; Pavelka, 2016). Recent literature has identified more than half the states in America have implemented

restorative justice programs (Umbreit & Armour, 2010; Pavelka, 2016) (see Figure 1). The prevalence of restorative justice practices reflects a broader desire to address crime from a more holistic perspective. Questions remain as to whether the increased prevalence of restorative justice is having an impact on different outcome measures of success (i.e., recidivism, stakeholder satisfaction, restitution compliance).

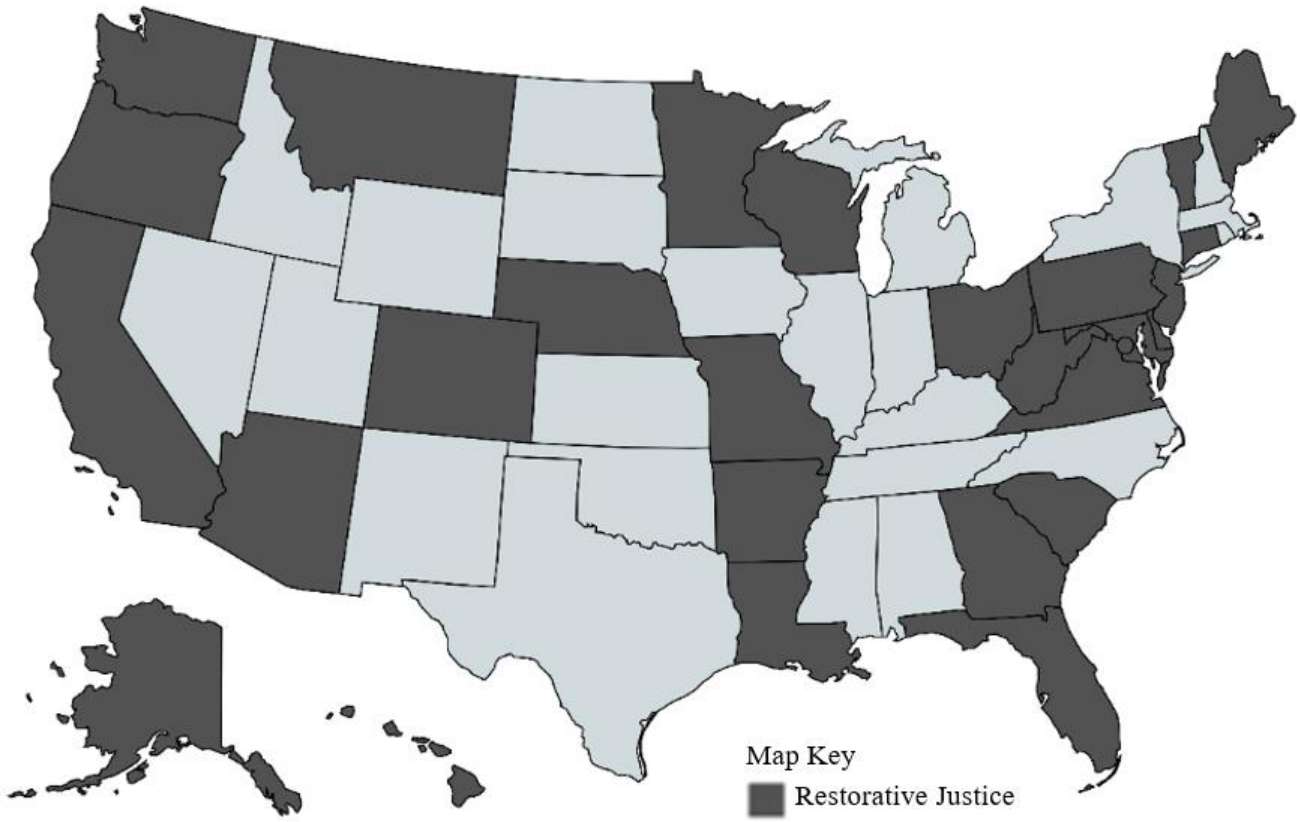

Figure 1. Presence of restorative justice in the United States. Data used from Umbeit & Armour, 2010 and map adapted from Pavelka, 2016

Unintended Consequences of Restorative Justice

Advocates for restorative justice stress the intended benefits of programs to stakeholders. Proponents believe successful restorative justice both repairs and prevents harm while providing cost-effective solutions for strengthening communities. More specifically, restorative justice programs help offender realize the true impact of their behavior and create opportunities for them

to work toward successful reintegrated into society. This process subsequently decreases recidivism and promotes public safety. Offenders may perceive restorative justice as a fairer and timely process compared to the traditional criminal justice system. Restorative justice empowers victims to voice their harms, confront their offenders, and to be involved in decision-making, all of which foster victim recovery and satisfaction. Despite the numerous intended benefits of restorative justice, skeptics have identified a host potential unintended consequences.

There are many cautions against the use of restorative justice practices within the United States. Objections to the use of restorative justice are based on ideological concerns as well as worries over practical limitations. America has a long legacy of enacting "feel good" policies which are well-conceived but poorly-realized (Miller, Gibson, & Byrd, 2008; Schiavone & Jeglic, 2009). The ideals of restorative justice and its express goals may be overly-ambitious and conceptually troubling to some scholars. Proponents against the practices of restorative justice debate its appropriateness and feasibility given gaps between program expectations, practical limitations, and feelings that there may be unintended consequences to stakeholders. Proponents against its use are skeptic of positive evaluations and questions about its utility and feasibility persist: *Are the goals of restorative justice unrealistic or overly ambitious? Is restorative justice appropriate for severe crime types? Is restorative justice cost-effective as a crime solution? What are the potential unintended consequences of restorative justice programs?* Though research addressing these specific questions is lacking, past literature has provided some merit for these concerns.

Dialogue-based restorative justice programs are not always feasible. There are a host of caveats and situations that make restorative justice difficult, inappropriate, or impossible. For stakeholders to effectively address harms through restorative dialogue several conditions must be

met: (1) participation must be voluntary, (2) held in a safe location, (3) facilitated by an objective party, (4) respectful dialogue, and (5) equal prioritization of stakeholder needs (Zehr, 2015). When any of these stipulations are not met, the restorative process is corrupted and subject to poor outcomes. This is especially true for voluntary participation and equal prioritization of stakeholder needs. The current study examined the voluntary nature of participation as a meaningful influence on success outcomes using distinct categories. Unlike offenders' sentences, or mandated, to restorative justice, when participation in restorative justice programing was at the choice of the offender and not contingent on other obligations or circumstances it was considered completely voluntary. A third category of "somewhat voluntary" was used to describe offenders are given the option to participate in restorative justice programs, but when their declining to do so or failure to complete said programs results in some form of legal consequence. The distinction of the somewhat voluntary category is important because offenders may be pressured or coerced into participation and therefore may be less engaged and cooperative. The cooperation of stakeholders as active participants in the justice seeking process is another defining characteristic of restorative justice (Zerh, 2015). Several factors may hinder cooperation between stakeholders. When offenders are mandated to participate in shared dialogue they may be disrespectful, unruly, or disengaged. The potentially negative experiences of victims of violent crime, domestic violence, sexual assault or abuse, or the survivors (i.e., family members) of victims of homicide may be further traumatized by there participation. Disengaged or confrontational offenders can be detrimental to victims' wellbeing. Moreover, unreceptive offenders may refuse to take accountability for their actions, be uncooperative or confrontational, or fail to meet victim expectations for remorse, all of which impedes emotional

recovery, particularly for sexual or violent crimes (Bibas & Bierschbach, 2004; Mika et al., 2004; Van Wormer, 2009).

Restorative justice is meant to be a balanced approach to responding to crime. In practice however, programs may be unbalanced, prioritizing the needs of one stakeholder over another. Lacking program fidelity often results in poor outcomes for stakeholders. Restorative justice programs are initiated by and organized around meeting offender needs first (i.e. rehabilitation), which may not be useful or compatible with the needs of the victim or community (Mika et al., 2004). When this occurs, the restorative justice practice is corrupted, and victims are less likely to experience any benefit for their participation while offenders receive leniency in terms of formal punishment for their offense. Conversely, unbalanced practices may be victim-oriented and shame offenders without means for redemption. These practices foster isolation and stigmatization of offenders instead of creating opportunities for their involvement. This effect may be traumatic for offenders due to the personal nature of restorative justice practices in comparison to more traditional forms of justice, which are less personal.

The stipulations of successful practice raise important concerns about the appropriateness of restorative justice use in specific types criminal cases. Prior research has scrutinized the appropriateness of restorative justice uses with adult, repeat, or serious criminals (Hudson, 2002; Shapland et al., 2008). Though restorative mediation and conferencing have been used in cases of serious crime (i.e., domestic violence/intimate partner violence, sexual assault) issues for victims have been identified. Yet, some scholars maintain that restorative interventions such as mediation are more meaningful for victims and more effective than the traditional criminal justice system at addressing sexual offenses symptomatic of a career sexual criminal (Koss, Bachar, & Hopkins, 2003). Financial restitution agreements alone do little to improve victims'

wellbeing or sense of security (Mika et al., 2004; Van Wormer, 2009). However, several evaluations failed to find positive impacts for mediation on reductions to offender recidivism or victim fear (Lane et al., 2005; Miers et al., 2001; Roy, 1993; Schneider, 1986; Shinar & Compton, 1995).

Though research offers strategies for avoiding these negative impacts, lack of standardized delivery and the unpredictability of stakeholder participation make implementing safeguards difficult (Weitekamp & Kerner, 2012). Additionally, it is difficult to predict the full spectrum of poor outcomes of newly implemented programs, no matter how well planned (McCold, 2003; Miller, Gibson, & Byrd, 2008). The literature that examines the implementation of restorative justice use with serious crime types is minimal, and findings of its effectiveness are mixed. Further research is needed to understand the full spectrum of impacts on involved stakeholder, both positive and negative (Braithwaite, 1999; Mika et al., 2004). However, without establishing evaluation credibility there will continue to be mixed support for programs and reservations about the use of restorative justice under any circumstance (Mika et al., 2004).

Limitations of Dependent Measures

Inconsistent Definitions of Success. Impact evaluations, also known as outcome evaluations, examine the effects of a treatment, in this case restorative justice programs, on participants. The recent popularization of restorative justice in the United States is generating increased interest about its effectiveness. The feasibility of future use of restorative justice hinges on our ability as researchers to accurately measure and report its various impacts on stakeholders. Gaining an understanding of how success is measured can go a long way in making informed suggestions for the standardization of outcome variables and improve the replicability of future research. How prior literature has operationalized success is influenced by several

limitations related to practice. The goals of restorative justice, reparation of harms (i.e., victim recovery), offender rehabilitation (e.g., recidivism, accountability), and restored stakeholder relationships, do not readily lend themselves to scientific measurement or identifying causal relationships (Presser & Van Voorhis, 2002; Umbreit,1989; Zehr, 2015).

Reparation of harms means different things to different people. While some victims suffered losses that can be undone with financial restitution, other harms may be more psychologically or emotionally bound (Johnstone, 2013; Pranis, 1997; Van Ness & Strong, 1997). Determining success in terms of achieving the former is relatively simple. *Was the victim paid financial restitution?* (Braithwaite, 1999; Presser & Van Voorhis, 2002; Roy, 1993; Umbreit, 1994). But *how does one go about measuring the latter concepts pertaining to psychological well-being or victims' emotional recovery?* (Umbreit,1997). Some evaluations have addressed this question by assessing victim satisfaction or perceived fairness with participation in restorative justice and with its outcomes (i.e., obligation agreements) (Hotaling & Buzawa, 2003; Umbreit, 1989, 1994a, 1994b). While this is not a direct measure of concepts like trauma or distress, it gives evaluators a general idea about what victims take away from the restorative justice process. Less research has attempted to directly measure psychological and emotional outcomes specifically. Studies have measured Post-Traumatic Stress Symptoms (PTSS), fear of victimization, anxiety, depression, and anger (Angel, 2005; Parsons & Bergin, 2010; Umbreit, Coates, & Kalanj, 1994). However, these studies also suffer from a lack of standardization in terms of consistency between specific measurement instruments. This issue is further compounded by the reliance of self-report data, the use of poorly crafted instruments, and the variable sensitivity of instruments measuring similar outcomes (Presser & Van Voorhis, 2002).

The most obvious, and perhaps most well-established, measure of offender rehabilitation is recidivism. To the extent that behavioral change encompasses the concept of offender rehabilitation, criminal desistence may be a valid measure. However, prior literature notes several issues with its use. First, recidivism may be a poor measure for offender rehabilitation because offenders may just be learning how to avoid detection and apprehension (Mair, 1991; Maltz, 1984). Conversely, offenders under the supervision of probation or parole, or any individuals with a criminal label or criminal history, have a greater likelihood for future contact with the formal criminal justice system (e.g., parole violations, re-arrest, court appearances) independent of crime commission (Bonta et al., 2006). The second issue with using recidivism as a proxy for offender rehabilitation is the lack of standardization with the way researchers employ this measure.

The obstacle of consolidating research using varying operational definitions of recidivism is not a problem which is unique to restorative justice research (Blumstein & Larson, 1971; Harris, Lockwood, Mengers, & Stoodley, 2011). Some scholars have addressed this issue by collapsing recidivism categories, which can result in confounding results. Whereas others choose one measurement of recidivism and exclude all other studies, effectively limiting their ability to generalize findings due to a reduced sampling frame. Finally, recidivism alone may not wholly capture what it means to be rehabilitated for all offenders. A host of alternative outcome measures have been employed by prior literature to address this shortcoming. Accountability, acknowledgement, expressions or remorse or guilt, verbal apology, written apology, and empathy are some variables used to capture other constructs of offender rehabilitation (Tangney, Stuewig, & Hafez, 2011). Because these outcomes are difficult to measure and are conceptualized by researchers in different ways, the obvious issues among this subsection of

research are problems with validity, reliability, and the fidelity in measurement. Additionally, expressive variables (i.e., remorse, empathy, guilt) may be disingenuous or a byproduct of offender cohesion (e.g., mandated apology letters, remands by to court based on failure to acknowledge wrong-doing) (Choi & Severson, 2009).

Restoration of stakeholder relationships (e.g., community connectiveness, community relations) is perhaps the most challenging restorative justice goal to evaluate. Primarily because community-level data, beyond crime rates, is difficult to acquire and impractical for analytics (Hartnagel, 1979). Additionally, restorative justice is an intimate process. Although the community is meant to be involved, especially in dialogue-based modalities, the benefits of restorative justice programs are not likely to have a measurable impact at the macro-level (Kurki, 2000; McCold, 1996). This is not because the experiences of stakeholder participation do not penetrate their respective communities, but rather any impacts at the community-level would not necessarily be apparent, or when measurable (e.g., collective efficacy, social cohesion) would violate the assumptions of a causal relationship (Jenson, 2010; Presser & Van Voorhis, 2002). Based on these factors, the use of community-driven success outcomes in restorative justice evaluations is improbable.

Research Question One. Given the breadth of possible outcome measures for determining success, the first research question of this investigation is exploratory: *How has success been defined in restorative justice evaluations within the United States?* The researcher predicts that: (1) a variety of dependent definitions of success, as well as different instruments measuring similar outcomes, will be abundant in the study sample, and (2) consistent with prior criminal justice and restorative justice literature, recidivism in its may operational forms, will

emerge as the most common dependent measure. Examining the ways in which past research has captured success will help to identify gaps between restorative justice goals and quantifiable outcomes. Additionally, examining the spectrum of outcome measures may draw attention to areas that need greater attention in terms of measuring the impact of program participation for different stakeholders. Over-reliance on offender-centric and, or victim-centric outcomes would indicate that more attention is needed to evaluate the communal impacts of restorative justice programs. The results of this inquiry have implications for future research efforts.

Limitations of Past Meta-Analyses

Research efforts to overcome methodological shortcomings of less rigorous program evaluations have given way to more systematic and comprehensive examinations. To better synthesize the body of evaluative restorative justice literature meta-analyses have been conducted to estimate summary treatment effect sizes for a variety of outcome measures including recidivism and satisfaction (Bradshaw et al., 2006; Latimer et al., 2005; Wilson et al., 2017). Though these methods have yielded more reliable, robust findings than single-studies or systematic reviews, support of the effectiveness of restorative justice is mixed and positive impacts have been small to moderate in size (Wilson et al., 2017). Despite increased empirical attention directed at examining program success, the results of extant meta-analyses are limited in their ability to inform debates about the effectiveness of restorative justice within the United States and have overlooked potential moderating effects (e.g., voluntary offender participation, methodological quality).

With few exceptions, the results of meta-analyses have favored of restorative justice effects to reduce recidivism relative to non-restorative treatments. Prior meta-analyses focusing

on various restorative justice modalities and different types of recidivism outcomes have found an average of 8-30% reduction to recidivism when compared to traditional justice processes (Bonta, Wallace-Capretta, & Rooney, 1998, 2002; Bradshaw & Roseborough, 2005; Latiemer, Dowden, & Muise, 2001, 2005; Nugent, Williams, & Umbreit, 2003, 2004). Similarly, meta-analyses examining re-offending specifically had also supported that exposure to restorative justice programs results in less recidivism than control groups (Hayes & Daly, 2004; Luke & Lind, 2002; Rodriguez 2005, 2007). Conversely, a couple of meta-analyses from the 1990s found negative treatment effects for offender recidivism whereby participation in restorative justice resulted in greater recidivism outcomes than control groups (Niemeyer & Shichor, 1996; Roy, 1993). The variation in meta-analytic design such as study identification techniques, types of restorative justice programs included within samples, and measurement of recidivism, as well as a lack of consideration for potential moderating variables, greatly limit the generalizability of these results.

The different programs of interest in these early investigations included restitution, community service (Bonta et al., 1998), mediation with only juvenile offenders (Nugent et al., 2003, 2004) or combined multiple programs but did not observe heterogeneity by program type (Brawdshaw & Roseborough, 2005; Latimer et al., 2001, 2005). More recently, meta-analyses have been conducted to directly compare variation in treatment effects by program delivery, offender age, race, gender, and crime severity (Bergseth & Bouffard, 2007, 2013; Hayes, 2005; Rodriguez, 2007; Sherman, Strang, & Woods). However, these research efforts failed to examine importance of various delivery differences such as voluntary participation, program type, utilized poorly-randomized samples or suffered from self-selection bias or other methodological shortcomings. Building on past research, meta-analyses should continue to investigate potential

differences in treatment effects based on program components and offender characteristics using studies of high methodological quality. There is a pressing need for continued meta-analytic research concerning the effectiveness of restorative justice programs in the United States. The necessity for continued research is driven by the limitations of prior evaluation efforts and a lack of understanding what, if anything, *works* in restorative justice and how future research and practice should proceed moving forward.

Existent meta-analyses cannot adequately address questions about the impact of restorative justice in the United States or inform evidence-based practice for several reasons. First, all previous meta-analyses have utilized mixed, international samples to generate pooled effect size coefficients. Therefore, inferences about the overall effectiveness of restorative justice in the United States cannot be drawn. Second, few meta-analyses have explored the variation in effect sizes due to delivery characteristics or program type. Specifically, past research has either focused on one form of restorative justice or has neglected to account for the variance due to program type. Improved methods are also needed to include variables such as offender age, or status of participation (i.e., voluntary vs. mandated). Third, operational definitions of success measures are limited. Though recidivism is a universally accepted outcome measure for program impact, little research has also attempted to measure success in ways that are intrinsic to the restorative justice model or to assess the validity of outcome measures in the context of restorative justice goals (e.g., client satisfaction, restitution compliance). Fourth, despite meta-analyses being a more sophisticated and robust statistical tool for quantifying treatment effects while controlling for variations in sampling size and statistical power, other methodological flaws from studies are carried over into their samples. These flaws are evident in the randomization of assignment to treatment and control groups, poor validity of instruments, and

failure to control for extraneous variables. Eliminating unsound methodological designs and controlling for the objective methodological quality within meta-analyses is critical for resolving these carryover limitations.

Research Question Two. *How rigorous are empirical evaluation designs?* The researcher hypothesized that objective methodological quality of impact evaluations would be negatively skewed with fewer evaluations of high quality (i.e., case-control, quasi-experimental, and experimental design). By examining pooled treatment estimates of program impact, the researcher gleaned general insights about the overall effectiveness of restorative justice as well as the moderating effects of program delivery characteristics. This meta-analytic investigation built on prior work in several meaningful ways and had direct implications for improved evaluative design and evidence-based practice. Comparing the objective strength of a representative sample of empirical studies helped to identify areas in methodological design that need improvement as well as gave context to the collective body of evaluative research within the United States.

Research Question Three. *How successful are restorative justice programs?* This meta-analysis used the outcome measures of recidivism, as measured by subsequent arrest, victim satisfaction, and restitution compliance as proxies for program success. Consistent with prior empirical research that supports positive effects of restorative justice on recidivism and victim satisfaction, the researcher predicted that the success across all three outcomes would be positive. The researcher also predicted that different program types (i.e., restitution, mediation, circles, conferencing), relative methodological strength (SMS score), as well as study characteristics such as offender age category, offense type category, and voluntary participation would account for heterogeneity within the sample (i.e., meta-analyses results). Specifically,

dialogue-based programs, such as mediation, conferencing and circles, with have greater positive impacts for stakeholders than restitution schemes, which tend to be less inclusive of stakeholder engagement in the restorative process.

Contributions of the Current Study

The obstacles of evaluating the success of restorative justice programs are considerable. Generally, complications are due to the complex nature of restorative justice goals and the difficulty with which outcome variables are defined or measured. Even when evaluative scholars agree on what constitutes success, evaluative design takes different forms with varying strengths and weaknesses. The need for further research is even greater when placed in the context of United States where the use of these programs is increasing as an alternative to traditional court processes and formal punishment. Not only are American-based evaluations scant in comparison to other countries but given the over-reliance of recidivism as an outcome measure for crime research in general, it is important to assess what, if any, non-recidivistic measures of success are used by in empirical research. The current study contributes to the growing evaluative body of research concerning restorative justice by addressing these concerns through unique study design.

Systematic searches identified a select sample of studies for coding and analysis. Only empirical studies form the United States were included in the sample. Success was measured using multiple outcome variables which more accurately and holistically reflect restorative justice goals (i.e., recidivism, stakeholder satisfaction, and restitution compliance). The threshold for methodological rigor in the current study was both rigorous and objective. Finally, studies were only eligible for inclusion if they receive the highest scores for methodological quality among case-control, quasi-experiments and experiments. The final sample of studies selected for

meta-analysis were used to generate summary treatment effects. Following meta-analyses for subsequent arrest, victim satisfaction, and restitution compliance outcomes, subgroup analysis was employed to assess heterogeneity in several meaningful ways. Subgroups examined the moderating effects of program type (i.e., restitution, mediation, circles, conferencing, hybrid, other) as well as delivery characteristics on subsequent arrest outcomes.

Past meta-analytic research has either compared success outcomes for one program type, or has included multiple program types without exploring the influence of treatment modality. By including five categories of programs and using subgroup analysis, the current study addresses these limitations. Additionally, meta-analyses are designed to give more weight effect-sizes yielded from more rigorous studies, such as studies with matched comparison groups versus non-matched groups or larger sample sizes versus smaller sample sizes. However, meta-analyses are limited in their ability to account for other methodological shortcomings such as those identified by the SMS (e.g., random assignment contamination, varying time points, etc.). Subgroup analysis comparing the summary treatment effects between subgroups of studies with different SMS scores will help detect variance due to objective methodological strength.

Restorative justice in the United States is primarily used with juvenile offenders and less serious crime (i.e., non-violent and property offenses). By comparing data on the success of programs which used adult offender or mixed age offender sample, the author can detect if these programs are better suited for juveniles or adults. Scholars have also expressed concern over the potential negative impacts of non-voluntary offender participation or use of restorative justice in more intimate, serious crime (i.e., violent and sex offenses) in restorative justice. The debate about the impact of type of offender participation is further complicated by the assertion that positive program impacts may be an effect of selection bias when offender participation in

restorative justice is voluntary. By comparing the voluntary nature of offender participation on multiple outcomes, the results of this study investigate the validity of these claims. Finally, results of subgroup analysis address concerns over effectiveness of restorative justice in cases of serious crime. The findings of the current study provide clarity about the success of restorative justice programs, further the field of evaluative research by suggesting methodological improvements moving forward, informs the debate about the utility of these programs within the United States, and suggests which programs and delivery characteristics are associated with greater success.

Methodology

Systematic Search of Literature

Capturing a representative sample of studies within an expansive field of literature is challenging. Inconsistent or flawed search methods in individual studies are subject to misinformation and the collection of biased samples. Systematic search protocols offer a scientific approach to collecting unbiased investigative samples from the larger body of literature on a given topic (Higgins & Green, 2008). As a component of the systematic review process, systematic searches are most commonly used in medical research to gather available evaluative data and used to summarize findings about interventions to inform best patient care practices (Bero et al., 1998; Cook, Mulrow, & Haynes, 1997; Cook & West, 2012). Consistent with recent research in the field of criminology, the current study utilized a systematic search protocol to identify representative evaluative research on restorative justice programs for the purposes of conducting a meta-analysis (Sherman, Strang, Mayo-Wilson, Woods, & Ariel, 2015). The processes of systematic searches follow rigid scientific designs to ensure continuity in future research to accurately replicate steps for arriving at a specific sample of studies from the broader literature (Petticrew & Roberts, 2008). Systematic searches employ clear language for search parameters based on the research agenda and detail the strategies used for source identification including specific search terminology, names of electronic libraries or databases used in the search, and how many sources searches yield (Higgins & Green, 2008; Wilson et al., 2017). The current study used a systematic search of restorative justice literature to identify eligible evaluation studies that will be subjected to a meta-analysis.

First, searches were carried out using specific terms. These search terms, or keywords, were chosen using the Participant/ Problem/ Population, Intervention, Comparison, Observation/ Outcome (PICO) method (Aromataris & Pearson, 2014; Santos, Pimenta, & Nobre, 2007). This

process involved generating lists of searchable words on each domain central to the research inquiries, including: (a) how is program success being measured?, (b) what is the methodological standard of impact evaluations?, and (c) how effective are restorative programs? With these questions in mind, the PICO method was replicated to compile all possible search terms to derive the desired sample of articles from relevant literature. The keywords selected were reflective of both broad constructs such as “restorative justice”, “evaluation,” or “success” as well as verbiage specific to individual program types (i.e., restitution, circles, mediation, conferencing) and research methodology (e.g., randomized control trial, quasi). Second, the search terms were organized using “AND”/ “OR” Boolean phrases to generate one comprehensive search string composed of 35 individual keywords (Montori, Wilczynski, Morgan, & Haynes, 2005). An overview of search parameters (i.e., keywords) and results are outlined in Table 1. Third, the search string was used across pre-selected electronic databases (Anders, & Evans, 2010; Bartels, 2013).

Table 1.

Systematic search notes

Date of search	February 11, 2019
Search string:	Search string: “restorative justice” OR “restorative” AND “mediation” OR “circles” OR “conferencing” OR “restitution” OR “reconciliation” OR “victim-offender mediation” OR “victim-offender conferencing” OR “victim-offender reconciliation” OR “victim offender dialogue” OR “victim-offender mediation” OR “restorative justice conferencing” OR “family group conferencing” OR “community group conferencing” OR “restorative group conferencing” OR “restorative justice circles” OR “repair of harm circles” OR “sentencing circles” OR “circles of support and accountability” OR “facilitated dialogue” OR “family unit meetings” OR “family group decision making conference” OR “healing circles” OR “sentencing circles” OR “peacemaking circles” OR “conflict resolution” AND “assessment” OR “evaluation” OR “impact evaluation” OR “effect” OR “effectiveness” OR “experiment” OR “quasi” OR “assessment”
Google Scholar results:	6,587
EBSCOhost results:	292
Gale Virtual Reference Library results:	14
WorldCat results:	77
IJRJ results:	23
Restorative Justice: An International Journal results:	17
Total results:	7,008

For the purposes of this investigation several search engines were employed to maximize the likelihood that both published and unpublished evaluations would be identified for screening into the sample. The protocol specified a total of four search engines including those most likely to house empirical, published research, EPSCOhost and Gale Virtual Reference Library, as well as more accessible search engines like Google Scholar and WorldCat, to capture unpublished investigations. Next, specific academic, peer-reviewed restorative justice journals were reviewed electronically following the same search protocol. These were the *International Journal of*

Restorative Justice (IJRJ) and *Restorative Justice: An International Journal*. However, the recent establishment of these journals, created in 2018 and 2013 respectively, coupled with the international nature of article publications, resulted in minimal additions to the initial sample. Finally, search results were exported into one reference list for consideration of which studies would be included in the sample in addition to using clear search criteria such as keywords and systematic selection of electronic databased, a set of unique steps was needed to address challenges with using these search engines.

The limitations of specific electronic database search tools were addressed in several ways. First, search engine keyword fields with character limits and/or Boolean phrase maximums were accounted for using "AND" phrases as points of separation between "OR" string phrases such that there were two lines of search parameters containing the same language as the single search string. This additional step of creating hard breaks within the search string, instead of relying solely on Boolean phrase separations yielded more results than including the single 35-word search string within on search tool field. Second, Google Scholar limited search returns to a maximum of 1,000 articles. In order to capture all available literature, a series of searches using the search string were carried out by publication year beginning with the first chronological publication in 1972 and ending with the year 2019.

Study Selection Procedure. Several inclusion criteria were used to refine the initial sample for coding. The eligibility criteria were consistent with the current study's working definition of restorative justice. To qualify as a restorative justice program evaluation, the intervention under scrutiny must have: (1) self-identify as "restorative justice", "restitution", "mediation", "conferencing", or "circles", and/or; (2) brought together two or more stakeholders of a crime (i.e., victim, offender, community member(s)) to; (a) talk about an offense or the

harms caused by crime, or; (b) to create agreed upon obligations (e.g., agreement or contract for earned redemption). In utilizing a broad, inclusive operational definition of restorative justice the researcher was able to acquire a sample that best represents the available body of evaluative research for consideration into the sample. These inclusion criteria also reflect the basic principles of restorative justice as defined by prior literature which helped to identify programs that utilized restorative elements but not restorative language (Braithwaite, 2004; Walgrave & Bazemore, 1999). Furthermore, there were no age, crime type, or criminal history restrictions for individual stakeholders for studies to be included in the sample. These variables were later used to examine the variance in treatment effects using moderator meta-regression analyses.

In addition to program characteristics eligibility criteria, studies must have evaluated restorative justice programs based within the United States. As discussed earlier, the strategic exclusion of studies which evaluated the impact of foreign restorative justice programs provides a narrower focus than prior research. The current study addresses a distinct gap within past literature by examining the utility of restorative justice programs juxtapose the critical issues that exist within the criminal justice system model of the United States. Only qualitative evaluations were eligible for coding. Finally, duplicate articles were removed from the sample. Then, additional sources were introduced into the sample from a review of bibliographies. These articles were also scrutinized using exclusion criteria and unfit studies were eliminated from the sample. A final coding sample of 121 studies was used for coding study outcome measures, study sample characteristics, and scored based on their methodological rigor (see Table 2).

Table 2.

Study selection processes for coding sample

Reasons for study removal	N
Search total (7,008) + bibliography study additions (327)[a]:	7,335
Studies removed for being outside the United States:	-6,714
Studies removed for being qualitative, non-comparative:	-412
Studies removed for not meeting definitional components:	-51
Studies removed for being duplications:	-158
Final coding sample:	= 121

a. 11 studies were not obtained in full text for review

Coding Procedure

A complex coding scheme was employed to systematically record study information (see Appendix A). To address the three research questions, data was coded for basic study characteristics (i.e., year of publication, type of publication), use of specific dependent measures captured as binary, and methodological design (i.e., methodological strength scores). Basic study characteristics such as author(s), type, and year of publication were captured along with potential moderating variables such as offender age group, voluntary offender participation, and program type. Outcome measures of program success were captured in the coding phase of this study. Dependent variable information was coded using binary choices (0 = no; 1 = yes) for pre-specified outcome measures drawn from prior research, such as recidivism outcomes (e.g., re-arrest, court petitions, reconvictions, incarcerations) or perceptions of fairness or satisfaction. Open-ended fields captured dependent variables not included within options of dichotomous outcome measure codes included in the coding manual. For instance, the original coding manual did not include an outcome for perception of negative labeling, which was a measure, which emerged from the current study sample. The coding manual was periodically updated with

outcomes not previously specified during the coding process. The results of success outcome coding were used to determine study eligibility for inclusion in meta-analyses.

The current study design employed an adapted version of the Maryland Scientific Methods Scale (SMS), to rank the relative strength of methods based on design characteristics (e.g., random assignment or selection, matched samples, pre-and-post testing). The SMS is a 5-point scale that outlines criteria for categorizing study design in terms of rigor (Farrington, Gottfredson, Sherman, & Welsh, 2002; Madeleno & Waights, 2010; Sherman et al., 1998). The SMS is a well-established global ranking system whereby a score of "1" represents the lowest possible methodological rigor (e.g., cross sectional correlation) and "5" represents the highest methodological strength score (e.g., randomized control trials) (Farrington et al., 2002). Whereas the original scale provides descriptions of methodological design components for scoring (see Table 3), the modified version of the SMS used herein includes more detailed criteria which helped more easily distinguish between method strength scores. The original SMS criteria for score adjustments of - 1 point value based on general methodological flaws (i.e., under-powdered sample size, inequivalent sample groups) have been criticized for their ambiguity, resulting scorer subjectivity.

Table 3.

Sherman et al. (1998) Maryland Scientific Methods Scale (SMS)

Score	Criteria
1	Correlation analysis, no control group, no attempt at establishing a counterfactual
2	Pre/Post comparisons, or a comparison group but without balancing or covariates
3	Difference-in-differences, balancing (OLS matching), but uncontrolled differences likely remain
4	Instrumental variable techniques or RDD, proper balancing (OLS matching), attrition discussed but not addressed
5	Randomized control trials, 'natural experiments', no selective sample attrition

Note: Source Farrington et al., 2002.

The current study follows Madeleno and Waights' (2010) SMS scoring guide, which contains stricter rules for adjusting raw scores -1 point value contingent on one or more violations of methodological design components unique to a particular test (see Appendix B). For instance, an evaluation using randomized control trials, SMS score category 5 (see Table 4), has three components for point deductions: (1) randomization is successful, (2) attrition carefully addressed or not an issue, and (3) contamination not an issue. Therefore, if any one of these three components are violated, the study would be a subject to a score adjustment of minus 1 point value, resulting in an adjusted SMS score of 4 (see Appendix B). Versions of the SMS scale have been used to evaluate the method quality in criminological research (Dodson, Cabage, & Klenowski, 2011; Fox & Shjarback, 2016; Lum, Koper, & Telep, 2011). However, the current study represents the first application of the SMS to assessing methodology rigor of restorative justice program evaluations. The SMS scores were subsequently used to select studies for inclusion in meta-analyses.

Table 4.

Madaleno & Waights (2010) Maryland Scientific Methods Scale (SMS)

Point Levels	Score Description
1	Either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. No use of control variables in statistical analysis to adjust for differences between treated and untreated groups or periods.
2	Use of adequate control variables and either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. In (a), control variables or matching techniques used to account for cross-sectional differences between treated and controls groups. In (b), control variables are used to account for before-and-after changes in macro level factors.
3	Comparison of outcomes in treated group after an intervention, with outcomes in the treated group before the intervention, and a comparison group used to provide a counterfactual (e.g. difference in difference). Justification given to choose of comparator group that is argued to be similar to the treatment group. Evidence presented on comparability of treatment and control groups. Techniques such as regression and (propensity score) matching may be used to adjust for difference between treated and untreated groups, but there are likely to be important unobserved differences remaining.
4	Quasi-randomness in treatment is exploited, so that it can be credibly held that treatment and control groups differ only in their exposure to the random allocation of treatment. This often entails the use of an instrument or discontinuity in treatment, the suitability of which should be adequately demonstrated and defended.
5	Reserved for research designs that involve explicit randomization into treatment and control groups, with Randomized Control Trials (RCTs) providing the definitive example. Extensive evidence provided on comparability of treatment and control groups, showing no significant differences in terms of levels or trends. Control variables may be used to adjust for treatment and control group differences, but this adjustment should not have a large impact on the main results. Attention paid to problems of selective attrition from randomly assigned groups, which is shown to be of negligible importance. There should be limited or, ideally, no occurrence of 'contamination' of the control group with the treatment.

Note: Source Madaleno & Waights (2010)

Additional sample characteristic and moderator variables were coded categorically for ease of comparison (see Appendix A). Age of offender was coded categorically as either juvenile

(i.e., less than 18 years of age) adult (i.e., at least 18 years-old) or mixed age groups when study samples included both juvenile and adult participants. Program type was catalogued into four broad domains: restitution, mediation, circles, or conferencing. During the coding process two additional program type categories emerged: “hybrid programs”, and “other” restorative programs. Hybrid programs, sometimes called balanced programs, employ different restorative programs based on case-level situational factors, but fail to control for program type within the evaluation. Hybrid evaluations assessed the effectiveness of restorative justice using a mixed sample of two or more programs (i.e., restitution, mediation, circles, conferencing). Whereas, the label of “other” was used to classify programs which employed treatment modalities which did not ascribe to the characteristics of traditional program types, but encompass some restorative elements (i.e., inclusion, engagement). Therefore, the category of “other” might include psycho-educational programs which utilize victim-impact panel designs.

For the purposes of subgroup analysis, offense type was captured as violent, non-violent, property, sex crime, or mixed when studies applied restorative justice to more than one crime category. Additional contextual information about studies was collected, such as specific offense type across 27 categories (e.g., arson, robbery, assault) offender criminal history, and specific stakeholder inclusion, but were ultimately excluded from moderator testing. To determine reliability of coding procedures, 25% of the articles (n = 32) were coded by a second independent coder and resulted in an overall agreement rate of 84.65% with kappa values ranging from k=.74-.94. This range was in line with standard mastery coding agreement (McHugh, 2012; Morse, Barrett, Mayan, Olson, & Spiers, 2002). Coder questions or discrepancies were used to further refine the coding manual with the author serving as the final decision for unresolved codes.

Meta-Analytic Strategy

Determining best practice for consolidating empirical evidence of treatment impacts is challenging. The current study conducted a series of meta-analyses to avoid the pitfalls associated with narrative synthesis literature reviews. Traditional, systematic reviews of empirical research rely on statistical significance (*p*) as a measure for detecting treatment effects. However, focusing on statistical significance is problematic for several reasons. First, sampling size heavily influences the ability to detect differences in mean values at the statistically significant level. Second, statistical significance as a measure for treatment effect is not versatile. Using *p* values may be helpful as an artificial abstraction to differentiate between means within an independent study, but the strength of statistical significance is not easily comparable between studies (Wilson, 2011). Review studies address inconsistency by providing thematic synthesis and observing the frequency of null hypothesis acceptance across studies. Although reviews employ expert, scholarly interpretation about the findings, reviews are fraught with subjectivity. Additionally, inferences made about the effect of a treatment fail to provide context for the size (i.e., magnitude) or direction of that effect (Wilson, 2011). Therefore, the effect of a treatment or intervention cannot be quantified by reviewing significance statistics alone (Bartolucci, & Hillegass, 2010; Egger, Dickersin, & Smith, 2001).

Meta-analyses prove helpful when a research question is met with equivocal findings between studies or there is an overabundance of literature to consider (Lipsey & Wilson, 2001). Combining data from multiple evaluative studies (i.e., pooling study populations) increases the power to detect differences between treatment and control groups. These analyses also control for sampling and standard error within each study to minimize the likelihood of special cause variation (Hedges & Olkin, 2014). The results of a meta-analysis describe the size and direction

of a treatment effects, as well as gage the consistency of those effect across studies. When the pattern of evidence across evaluations is not consistent, moderator testing explores the relationship between select indicator variables, such as study characteristics, and the generated effects (Lipsey & Wilson, 2001; Wilson, 2011, 2013).

The current study sought to provide clarity as to the impact of restorative justice programs on recidivism (i.e., subsequent arrest), victim satisfaction, and restitution compliance relative to traditional criminal justice. The pooled raw data from individual evaluation studies were used to calculate standardized effect size coefficients, mean confidence intervals, and generate overall effect statistics for each outcome variable. The general method involved using inverse variance weights to assign greater value to Log Odds Ratios (LORs) point-estimates calculated from the data of more precise studies using larger samples. A random effects model was used in favor of a fixed effects model because the author assumed heterogeneity in sample demographics. Although using mixed effects modeling results in penalized confidence intervals, this strategy was most appropriate for calculating summary effects given the use of populations that differ characteristically (Hedges & Vevea, 1998). All meta-analyses, heterogeneity testing (i.e., subgroup analysis), and publication bias testing were performed using Comprehensive Meta-Analysis Software, Version 3 (CMA) (Borenstein et al., 2018). Extracted data (i.e., sample sizes, dichotomous event outcomes) from a collection of eligible studies (N= 35) generated 43 effect size estimates and three summary treatment effects.

Sample. Not all articles eligible for methodological strength and dependent variable coding (N = 121) were included within the meta-analysis sample (N = 35) (see Table 5). A subset of studies was selected by a thorough review of study methods (i.e., analytic strategy), methodological strength, and outcome measures. To be included in meta-analysis studies needed

to have employed a design which compared restorative justice participation outcomes with non-restorative programs or practices. Studies were also removed from the meta-analysis sample if they did not contain enough data for calculating LORs effect size coefficients or scored below a three in methodological strength (e.g., single-case, cohort, or seriously flawed methodology). Finally, studies were included for meta-analysis only if they examined one of the three modal outcomes in the sample: recidivism, as measured by subsequent arrest, victim satisfaction, or restitution compliance (i.e., monetary agreements).

Table 5.

Study Selection Processes for Meta-Analysis Sample

Reasons for excluding studies	N
Initial sample derived from coding:	121
Studies without control comparison groups removed:	-58
Studies with SMS method scores 1-2 removed[a]:	-5
Studies with non-relevant outcome measures removed[b]:	-17
Studies without sufficient data for calculating LORs estimates were removed[c,d]:	-6
Final Meta-Analysis sample:	= 35

a. Although SMS category 2 studies include control groups, lower SMS scores were excluded from meta-analytics by design
b. Meta-analysis outcomes included recidivism (re-arrest), victim satisfaction, and restitution compliance
c. Studies which reported comparative data for some but not all outcomes were excluded
d. Studies reporting only continuous outcome data were excluded

The decision to use a narrow operational definition for the domains of recidivism, satisfaction, and type of restitution compliance was made for ease of comparison as well as limitations within the current sample. Although the recidivism meta-analysis could have included diverse measures (e.g., arrests, convictions, incarceration), the number of studies which assessed non-arrest outcomes and which also met other inclusion criteria were too few to assess variance

in treatment effects due to these differences (n = 4). Similarly, studies examining non-victim stakeholder satisfaction and non-monetary forms of restitution were scant (n = 1 and n = 2 respectively). Eighty-two studies did not meet these criteria and were eliminated from the sample leaving a final sample size of 35 studies for meta-analyses that accounted for 43 odds ratio effect size coefficients used to generate summary effects for subsequent arrest, victim satisfaction, and restitution compliance.

Sample Characteristics. The final sample of 35 studies, generating 43 independent LORs effect estimates present data from 10,627 treatment group offenders and 10,020 control group offenders. Table 6 provides the characteristics of the studies included in the analyses. The publication year range for studies was 1977-2018, with the majority (n = 24) of the studies published between 1990 and 2009. More than half of the sample was comprised of peer-reviewed studies from academic journal publications (62.85%), while non-peer reviewed studies made up about one-third of the sample (37.14%) and included book chapters (n = 5), theses or dissertations (n = 4), or technical reports (n = 4). The most common research design within the sample was experimental studies (42.86%), followed-up quasi-experimental (31.42%), and case-control studies (25.71%). Several studies within the sample reported on more than one outcome of interest (n = 8). Overall, subsequent arrest was the most common reported outcome measure (n = 31, 88.57%). Victim satisfaction and restitution compliance were each examined in 6 studies. Follow-up periods varied greatly between studies in the sample ranging from 4 months to 8 years. Eight studies did not include information on the length of follow-up. The size of treatment groups also varied considerably between studies, with the smallest sample size for subsequent arrest at 25, victim satisfaction at 26, and restitution compliance at just 14. The highest number of participants in a given study for subsequent arrest was over four-thousand (n =

4,189), whereas the largest treatment sample for victim satisfaction and restitution compliance outcomes were 259 and 162 respectively.

Table 6.

Meta-Analysis Sample Characteristics

Study Characteristic (n = 35)	Attribute categories	n	%
Publication year			
	1970-1989	6	17.14
	1990-2009	24	68.57
	2010-2019	5	14.28
Publication review status			
	Peer-reviewed	22	62.85
	Non-peer-reviewed	13	37.14
Publication type			
	Academic journal article	22	62.85
	Book chapter	5	14.28
	Report	4	11.42
	Thesis or dissertation	4	11.42
Research design			
	Case-control	9	25.71
	Quasi-experimental	11	31.42
	Experimental	15	42.86
Outcome[a]			
	Subsequent arrest	31	88.57
	Victim Satisfaction	6	17.14
	Restitution compliance	6	17.14
Treatment group size			
	Less than 100	15	42.85
	≥100 < 1000	18	51.42
	1000+	2	5.71
Follow-up period[b]			
	Less than 12 months	24	11.42
	≥12 months <24 months	13	37.14
	≥24 months <36 months	5	14.29
	≥36 months <48 months	2	5.71
	≥48 months <96 months	1	2.86
	96 months+	2	5.71

a. 8 studies reported on multiple outcomes
b. 8 studies were missing information on length of follow-up

Inverse Variance Weight Design. The first step in the analytic strategy was to choose the appropriate type of meta-analysis for the data. The current study utilized a mixed effects,

inverse variance weighted meta-analysis. Determining this type of meta-analysis was based on several assumptions and statistical objectives. The level of precision for each study effect size varies. Because some effect sizes are more accurate than others in terms of Independent and Identically Distributed Data (IID), standard computations (e.g., regression, ANOVA) cannot be used for statistical comparison (Egger, Smith, & Phillips, 1997; Hedges & Olkin, 2014). Meta-analyses address these issues by assigning more weight to studies with precise estimates and less weight to less precise studies (Wilson, 2011). The challenge then becomes what metric to use as a proxy for weighting estimates. Although sampling size could be used to determine estimate weights there are more accurate means for standardizing coefficients (Wilson, 2013). Standard error is a direct measure of precision. However, weighing estimates this way would result in studies with the least standard error (i.e., the most precise) having the least weight. Instead, the *inverse* of the standard error (i.e., squared standard error) determines the true estimate coefficient weight. This type of method is called an inverse variance weighted meta-analysis and it provides a statistical basis for standardized error of the effect size, confidence intervals, and homogeneity testing (Hedges & Olkin, 2014; Wilson, 2011).

Several sample characteristics inhibit the use of a fixed effects model meta-analysis. Fixed-effects models are used to estimate one true, underlying effect for a common population across multiple studies. Fixed effects models assume a value of *true zero* whereby variation is explained only by random error without room for other moderating factors (Field, 2001). However, these assumptions were not plausible for the current study which examined data that were accumulated from a series of studies that had been performed by different researchers using different methodologies. The program interventions differed across studies in ways that likely impacted the results of evaluations included in the sample. Because these studies were not

functionally equivalent, a common population or effect size cannot be assumed (Hedges & Olkin, 2014). Instead, random effects models assumed that each study did not share a common population, rather individual populations with potential for extraneous influence on effect size estimates (see Figure 2) (Bigby, 2014). As shown in Figure 2, the fixed effects meta-analysis model on the left displays a single grey parabolic arch, representing one population distribution, which house all study populations, displayed in color. While the random-effects meta-analysis model graph on the right depicts three separate population and distribution parabolas for individual studies. Therefore, a random-effects model was employed to estimate the mean of a distribution of effects across restorative justice evaluations assumed to have distinct populations too dissimilar to treat as homogeneous.

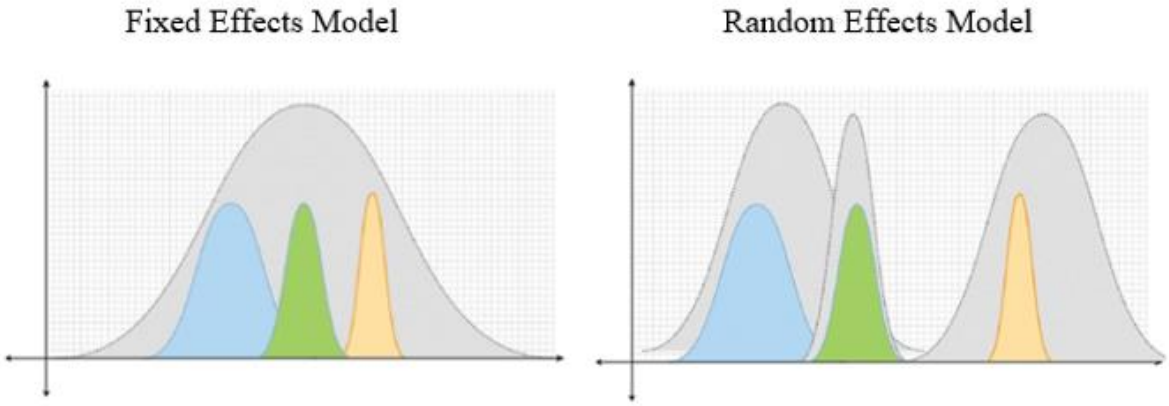

Figure 2. Fixed effects vs. random effects population comparison. source, (Bigby, 2014).

Calculating Effect Sizes. Following the coding phase of this investigation, data were extracted from a sample of 35 articles. First, treatment and control group sizes. Then, event counts were recorded for each group from each study. Replicating prior meta-analyses, these data were used to calculating odds ratio point estimates for the outcomes of subsequent arrest, victim satisfaction, and restitution compliance (Field & Gillett, 2010; Higgins & Thomas, 2002; Lipsey & David, 2001; Lipsey & Wilson, 2001; Thomas & Higgins, 2002). Odds ratio values are

designated for comparing two groups on a binary dependent variable (see Equation 1). A total of 43 effect sizes were generated across 35 studies. These included 31 subsequent arrest point estimates, 6 victim satisfaction point estimates, 6 restitution compliance estimate, and 1 summary treatment effect for each outcome variable.

$$ES_{sm} = In\left(\frac{ad}{bc}\right)\frac{\sqrt{3}}{\pi} \quad (1)$$

Estimate Corrections and Standard Error. Unlike standardized mean difference or other continuous outcomes, odds ratio statistics do not have standard error. To calculate standard error for these outcomes the asymmetry of the distribution of coefficients was corrected using log transformations (see equation 2) (Wilson, 2011). Corrected LORs were then used to calculate the standard error for weighing (see Equations 3) (Wilson, 2011).

$$ES_{In(OR)} = \log(ES_{OR}) \quad (2)$$

$$se_{OR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \quad (3)$$

Meta-Analysis Weighting and Confidence Intervals. A series of calculations were conducted in order to arrive at an inverse variance weighted value. First the general inverse variance weight was calculated using the corrected standard error values (see Equation 4). Using these estimated weights, a fixed-effects inverse variance weighted mean effect size was generated (see Equation 5). Although the current study specifies a random-effects model, this step is critical for obtaining the Tau (τ) statistic needed for subsequent analyses (Field & Gillett, 2010). Next, the standard error of the mean effect size was determined (see Equation 6) before confidence intervals were drawn around the standardized means (see Equation 7). Finally, the Z statistic was specified to the summary effect for interpretation of the variance between sample estimates (see Equation 8).

$$w = \frac{1}{se^2} \quad (4)$$

$$\overline{ES} = \frac{\sum w_1 ES_1}{\sum wi} \quad (5)$$

$$se_{\overline{ES}} = \frac{1}{\sum wi} \quad (6)$$

$$ES_{upper} = \overline{ES} + se_{\overline{ES}}1.96 \qquad ES_{lower} = \overline{ES} - se_{\overline{ES}}1.96 \quad (7)$$

$$Z = \frac{SE}{se_{\overline{ES}}} \quad (8)$$

Heterogeneity Statistics. When there is moderate heterogeneity within a meta-analysis sample the summary estimate may not be an accurate representation or oversimplification of the treatment effect (Thompson & Higgins, 2002; Wilson, 2011). There were two ways this investigation assessed heterogeneity. First, I^2 was calculated to detect variability within the sample not due to chance (see Equation 9). The *I* statistic is a simple test of inconsistency and describes the percent variance within the sample due to heterogeneity (Higgins & Thompson, 2002; Higgins et al., 2003). Common cause variation is a term used to describe variance in a sample due to chance. I^2 can be easily interpreted and unlike other heterogeneity statistics (i.e., Q) is not dependent on meta-analysis sample size. The thresholds or 'cut-offs' for interpreting the level of heterogeneity for I^2 are as follows: values below .5 (50%) are considered low heterogeneity, .5-.75 (50-75%) of heterogeneity is moderate, and .75 (75%) and above is considered high sample heterogeneity (Higgins & Thompson, 2002; Wilson, 2011). When high heterogeneity is observed in the sample, additional tests are required to examine the variation not due to chance.

$$I^2 = 100\% \; x \; \frac{Q-df}{df} \quad (9)$$

The second method for assessing heterogeneity uses a weighted sums-of-squared homogeneity of variance statistic expressed as *Q* (see Equation 10) (Higgins & Green, 2008). Cochran's *Q* tested how homogenous the sample was, or how much the sample lacked variance.

If *Q* is statistical significance at the p < .05 level, then heterogeneity is detected in the sample. The main limitation of this assessment the influence of sample-size on the power of the test whereby a too small sample sizes of studies leads to under-powered analysis or too much power when the number of studies included in the sample is large (Gavaghan et al., 2000; Higgins et al., 2003). Nevertheless, the *Q* statistic was included in this meta-analysis function as part of the random effects pooling method (DerSimonian & Laird, 1985; Hedges & Olkin, 2014). *Q* is required to calculate the random effects variance component, Tau squared (τ^2) (see Equation 11). Tau squared was then used to re-compute inverse variance weights for the random-effects inverse variance weighted meta-analysis (see Equation 12).

$$Q = \sum_{wi} ES_{i^2} - \frac{(\sum_{wi} ES_i)^2}{\sum wi} \quad (10)$$

$$\tau^2 = \frac{Q - df_Q}{\sum wi - \frac{\sum wi^2}{\sum w_i}} \quad (11)$$

$$w_i = \frac{1}{se_{i^2} + \tau^2} \quad (12)$$

Subgroup Analyses. Moderator, or heterogeneity, test used to account for treatment effect variance in the sample not due to chance. Because outcome variables were measured dichotomously, subgroup meta-analysis was used in favor of meta-regression analysis, which requires continuous variable measurement. Subgroup analyses in the current study were used to test several hypotheses concerning the potential influence program type, methodological strength, voluntary offender participation, offender age, and offense type have on subsequent arrest outcomes. Table 7 presents moderator analysis sample characteristics. In regard to the composition of program type, mediation evaluations accounted for most of the studies in the sample (n = 9, 25.71%), followed by conferencing (n = 7, 20.00%), other (n = 7, 20.00%), hybrid (n = 6, 17.14%), restitution (n = 4, 11.42%), and circle programs (n = 2, 5.71%). Only studies

with SMS scores 3-5 were included in the study sample. Of these 35 studies, 16 (45.1%) achieved a score of 4, followed by studies with an SMS score of 3 (n = 10, 28.57%). Studies with the more rigorous methodological design (i.e., SMS score of 5) accounted for about a quarter of the sample (n = 9, 25.71%). More than three-quarters of treatment group offenders were juveniles (80.00%) and no study within subgroup analysis included mixed juvenile, adult offender populations. The modal category of offender participation type was somewhat voluntary (n = 14, 40%), followed by mandated participation (n = 9, 25.71%), and voluntary participation (n = 7, 20%). Finally, more than half of studies reporting information on crime type were inclusive of both violent and non-violent offenses (n = 19, 54.28%).

Table 7.

Moderator subgroup analysis sample characteristics

Study Characteristic		N	%
Program modality (*n* = 35)			
	Restitution	4	11.42
	Mediation	9	25.71
	Circles	2	5.71
	Conferencing	7	20.00
	Hybrid	6	17.14
	Other	7	20.00
SMS score (*n* = 35)			
	3	10	28.57
	4	16	45.71
	5	9	25.71
Offender age type (*n* = 35)			
	Juvenile only	28	80.00
	Adult only	7	20.00
Crime type (*n* = 28)[a]			
	Non-violent, property	6	17.14
	Violent	1	2.86
	Sex crime	1	2.86
	Mixed	19	54.28
Voluntary participation (*n* = 30)[b]			
	Voluntary	7	20.00
	Somewhat voluntary	14	40.00
	Mandated	9	25.71

a. 7 studies did not specify offense type
b. 5 studies did not specify nature of offender participation

Publication Bias Testing. The final step in the analysis process was assessing for publication bias present in the sample. When publication bias is present, the results of a meta-analysis are misleading. Funnel plots, a type of scatterplot, were used following each meta-analysis to visually check for the existence of publication bias (Hedges & Olkin, 2014; Wilson, 2011). Studies with high precision (i.e., less standard error) and little heterogeneity were expected to fall near the average with low-precision studies dispersed evenly on either side of the average, creating an inverted funnel shape. When the distribution of studies deviates from the inverted funnel shape there is likely publication bias within the sample, suggesting

limitations for meta-analytic findings (Sterne & Egger, 2001). Formal tests for bias (i.e., common cause variation, special cause variation) were needed to quantify asymmetry in the distribution of plots.

Egger's regression intercept, Begg and Mazumdar rank correlation, and Duval and Tweedie's trim and fill further scrutinized for publication bias within the series of meta-analyses. Egger's test simulates a Y intercept equal to zero from a linear regression model for standardized effect estimates. Asymmetry, or discordance, of a funnel plot is assessed by dividing the normalized effect by the reciprocal of standard error (i.e., precision) (Egger, Smith, Schneider, & Minder, 1997). Expressed as Kendall's Tau, the results of Begg and Mazumdar's rank correlation test is used to examine the relationship between effect sizes and their variances (Begg & Mazumdar, 1994). Both Egger's test and rank correlation are limited in their accuracy for determining publication bias with under-powered samples, especially when used for meta-analyses with fewer than 25 studies. Trim and fill is a method used to adjust for the presences of publication bias. Trim and fill estimates the number of missing studies that could be missing from a meta-analysis. If the overall summary effect and corresponding confidence intervals are significantly improved after imputing missing studies, publication bias was an issue for the meta-analysis (Duval, 2005; Duval & Tweedie, 2000a, 2000b). Though statistical tests for determining publication bias have inherent limitations, taken together with LOR distribution on funnel plots these tests provide context for study results.

Results

Outcome Measures of Success

The current study thematically coded a sample of 121 evaluations of restorative justice programs to ascertain how "success" was operationalized within empirical research. These studies were compiled through systematic search procedures to ensure they accurately represented the available literature on impact assessments for restorative justice programs within the United States. The majority of the 121 studies evaluated mediation (n = 37, 30.58%) whereas circle programs made up the smallest proportion of the sample (n = 10, 8.26%) (see Table 8). Of particular importance to this study was the nature of dependent measures relative to the ascribed goals of restorative justice (i.e., repairing harms, offender accountability, and restoration of relationships) as compared to the traditional criminal justice system (i.e., deterrence). Past literature has noted the lack of standardization and reliance of recidivism outcomes across the dependent measures of restorative justice impact evaluations (Latimer et al., 2001; Presser & Van Voorhis, 2002). Furthermore, the use of recidivism as a proxy for effectiveness is heavily criticized due to its diverse application to the desistence of different types of behavior (e.g., re-arrest, re-conviction, re-incarceration) which entail different levels of contact with formal criminal justice with varying degrees of credibility (Blumstein & Larson, 1971; Harris et al., 2011; Maltz, 1984). Restorative justice fosters the involvement of victims, and communities, toward a more holistic redressing of the harms caused by crime than the traditional, retributive based, criminal justice system. Because focusing solely on offender-oriented outcomes constrains the ability to gain an understanding of the true impact of these programs on various stakeholders, the current study examines multiple outcomes (i.e., repairing harms, offender accountability, and restoration of relationships).

Table 8.

Composition of sample by program type

Program Type	n	%
Circle	10	8.26
Mediation	37	30.58
Conferencing	19	15.70
Restitution	19	15.70
Hybrid	16	13.22
Other	20	16.53
Total	121	100.00

Each outcome variable of evaluations was coded for individually. Therefore, the outcome measures of studies with more than one dependent measure were accounted for separately. The results of the current study support the predominant use of recidivism as measure of program success (see Table 9). The vast majority of the sample (97.5%) employed at least one measure of offender recidivism. These studies operationalized recidivism in different ways, most often relying on official record data. More than one-half of the sample operationalized recidivism as various forms of contact with the formal justice system (51.2%). Subsequent or re-arrest emerged as the modal outcome with about half of the sample (44.6%) measuring re-offending in this way. Subsequent court referrals or petitions were occasionally used to measure recidivism (18.2%). Other measures of recidivism, such as subsequent conviction, incarceration, or less formal measures, such as self-reported delinquency or disciplinary referrals, accounted for fewer outcome measures (see Table 9). Several studies reported on multiple outcomes. Frequency counts were based on the presence of each independent variable within a study.

Another popular success outcome was satisfaction (see Table 9). Of the 121 studies, 80 (66.1%) utilized at least one measure of satisfaction. Victim satisfaction was most frequent within the sample (30.6%) followed closely by offender satisfaction (24.8%). Fewer studies

assessed effectiveness by other stakeholder satisfaction (i.e., support stakeholders, citizen/community members). Over one-third of studies (36.4%) measured some form of restitution compliance. Monetary restitution fulfillment was more common than community service restitution compliance (23.3% and 14% respectively). The three domains of recidivism, satisfaction, and restitution compliance were the central focus of analysis in the current study. However, other measures of success were also present in the sample (see Table 10).

Table 9.

Most common outcome measures among the 121 restorative justice program evaluations

Variable Categories	n	%
Recidivism	118	97.5
Subsequent / re-arrest	54	44.6
Subsequent court referral / petition	22	18.2
Disciplinary referrals / expulsion / suspension	15	12.4
Self-reported delinquency / crime	8	6.6
Days until new first contact	6	5.0
Subsequent conviction / re-conviction	4	3.3
Subsequent offense severity	4	3.3
Risk of recidivism	2	1.7
Calls to police	1	0.8
Probation / parole violation	1	0.8
Subsequent incarceration / re-incarceration	1	0.8
Satisfaction	80	66.1
Victim / survivor satisfaction	37	30.6
Offender satisfaction	30	24.8
Support stakeholder (friend, parent, peer) satisfaction	7	5.8
Citizen/community stakeholder satisfaction	6	5.0
Restitution Compliance	44	36.4
Monetary restitution compliance	27	22.3
Community service restitution compliance	17	14.0
Other outcomes	60	49.9
Offender perception of fairness	15	12.4
Victim perceptions of fairness	15	12.4

Table 9. (continued)

Restorative contract / agreement reached	11	9.1
Restorative justice program completion	11	9.1
Perceived value in restorative process	8	6.6

Other operational measurements of success were diverse, though less prevalent than recidivism, satisfaction, and restitution compliance (see Table 10). About 10% of the overall sample focused on success measured as perception of fairness, program completion, or whether or not an agreement (i.e., contract) was reached between stakeholders. Less than 5% of studies accounted for more than 40 different outcomes. Most of these dependent variables of success were offender-centric (n = 24) and largely dealt with social, emotional, and behavioral outcomes. Social measures were concerned with the quality of relationships and offender attitudes toward relationships. These included offender pro-social relationships and attitudes, perceived social support, social competence, and relational aggression as well as attachments to parents and neighborhoods. Several success outcomes pertained to offender emotions, such as emotional regulation, empathy, shame, and guilt. Other types of dependent measures concerned offender self-concept (i.e., identity change, self-esteem, perception of negative labeling), offender scholastics (i.e., academic performance, attendance, commitment to school), re-entry (i.e., employment, re-integration success), or qualities of character (i.e., rebelliousness, honesty, responsibility, awareness).

Conversely, victim-centric measures of success were typically rooted in emotions such as fear, anxiety, anger, and hope (see Table 10). Evaluations also employed experiential outcomes like sense of involvement, perceived repair of harms, and how much of an opportunity victims felt they had to express their views. Least commonly, studies examined victim mental health (i.e., Post-Traumatic Stress Symptoms, anxiety, etc.) and general well-being. Finally, other

outcomes were related to support or citizen stakeholder experiences (e.g., community member/parent sense of involvement/engagement, staff perceptions), or included broader, pragmatic measures of success (i.e., increased frequency of diversions to restorative justice, cost comparison, change in school climate).

Table 10.

Least common outcome measures among the 121 restorative justice program evaluations

Variable categories	n	%
Social competence/ social skills	6	5.0
Academic achievement / performance	5	4.1
Offender accountability /responsibility	5	4.1
Offender apology / apology letter	5	4.1
Repair / restoration of relationships	5	4.1
School attendance record	5	4.1
Sense of community involvement	5	4.1
Number of diversions	4	3.3
Pro-social attitudes / peer relationships	4	3.3
Cost comparison (net)	3	2.5
Offender empathy	3	2.5
Perceived effectiveness of restorative process	3	2.5
Change in school climate	2	1.7
Cost comparison (per individual / case)	2	1.7
Family cohesion / attachment to parent(s)	2	1.7
Fear of re-victimization	2	1.7
Offender guilt	2	1.7
Offender parole success scale / re-integration success	2	1.7
Offender shame	2	1.7
Peer aggression / social, relational aggression	2	1.7
Perceived repair of harms / meeting of victim needs	2	1.7
Victim anger	2	1.7
Commitment to school	1	0.8
Improved victim feelings toward offender	1	0.8
Length of employment time after re-entry	1	0.8
Offender attachment to neighborhood	1	0.8
Offender awareness/knowledge of harms	1	0.8
Offender emotional regulation	1	0.8
Offender honesty	1	0.8
Offender identity change / positive self-esteem	1	0.8

Table 10. (continued)

Offender parent opportunity to express their view	1	0.8
Offender parent sense of involvement	1	0.8
Offender perceived level of social support	1	0.8
Perception of Labeling/ negative self-concept	1	0.8
Rebelliousness	1	0.8
Self-reported drug use	1	0.8
Sense of stakeholder engagement	1	0.8
Staff perceptions that they are helping stakeholders	1	0.8
Victim anxiety	1	0.8
Victim mental health / well-being	1	0.8
Victim opportunity to express their view	1	0.8
Victim sense of hope	1	0.8
Victim sense of involvement	1	0.8
Victim symptoms of post-traumatic stress syndrome	1	0.8

The first objective of this investigation was to determine how programmatic success was being measured in impact evaluations of restorative justice. It was hypothesized that a variety of operational definitions of success would be employed by prior research. Results from the dependent variable coding process support this hypothesis. A total of sixty-eight different outcome variables were identified in the current study sample of 121 studies. It was further hypothesized that recidivism would be used most often as a measure of success in program evaluations. This hypothesis was also validated given that nearly all (97.5%) of the studies included at least one form of recidivism as an outcome measure, with subsequent arrest having been used most frequently as a proxy of recidivism (see Table 9). Interestingly, a diverse set of dependent measures were identified in the sample but accounted for only a small proportion of studies (see Table 10).

Methodological Strength Scores

The second objective of the current study was to assess the relative quality of evaluation study design. Each of the 121 studies in the sample were scored using SMS criteria. Extended

scoring guidelines were used to adjust raw scores by deducting points as needed for methodological shortcomings (see Appendix B) (Madaleno & Waights, 2010). Appendix C contains an annotated table of individual study SMS score results. Overall, methodological rigor scores were well distributed among sample studies (see Table 11). Notably, nearly one third of the studies (n = 37, 30.57%) achieved the lowest possible SMS score. A similar proportion of studies fell within the middle three rank categories of 2 (n = 22, 18.18%), 3 (n = 29, 23.97%), and 4 (n = 22, 18.18%). Only 11 studies achieved the highest score for methodological rigor (9.09%). The scores of twenty-one studies (17.35%) required adjustment due to methodological defects. Seventeen score adjustments were made to studies having only one violation and four studies violated two or more SMS assumptions (see Appendix B). The most common violations among studies with adjustments dealt with high sample attrition (n = 7), variations in timing of intervention (n = 5), and contamination of randomized assignment (n = 4). Other studies suffered from issues with accuracy of randomization (n = 1), discontinuity in treatment delivery (n = 1), poor matching criteria for selection (n = 1), and unbalanced treatment and control groups (n = 1).

Table 11.

SMS Scores among the 121 restorative justice program evaluations

SMS Score Category	Raw Score n[a]	%	Adjusted Score[b]	%
1	37	30.57	37	30.57
2	18	14.87	22	18.18
3	27	22.31	29	23.97
4	17	14.05	22	18.18
5	22	18.18	11	9.09
TOTAL	121	100.00	121	100.00

a. Scored studies based on the SMS scoring tool, Source, (Madaleno & Waights, 2010)
b. Raw score adjustments of 1 point category deductions applied to the raw SMS score of studies with methodological shortcomings with violate 1 or more scoring criteria (see Appendix B).

The second research objective of the current investigation was to determine the objective quality of impact evaluations. It was expected that SMS scores would be negatively skewed such that a smaller proportion of studies would be of high quality. However, the breakdown of adjusted SMS scores reveals that only about half of the sample was comprised of studies with poor methodological quality (SMS scores 1 and 2). The other 51.24% of studies having received an adjusted SMS score between 3 and 5 employed case-control, quasi-experiment, and experimental designs. While some of these studies may have exhibited methodological shortcomings, they were not substantial enough to warrant an adjustment in score into SMS categories 1 and 2 (see Table 11). This distinction is important because it has implications for moderator tests assessing between study heterogeneity due differences in methodological strength. The proportion of studies with the highest score of 5 would have been doubled had score adjustments not been necessary, suggesting that the objective quality of randomized control trials within the sample varied greatly and that making inferences about the rigor of studies based solely on design is not reliable.

Meta-Analytic Results

The final research objective of the current study was focused on "taking stock" of the successful or restorative justice programs in the United States using a series of meta-analyses. The meta-analytic results for outcomes on offender subsequent arrest (i.e., recidivism), victim satisfaction, and restitution compliance are presented using forest plots. Forest plots provide a visual breakdown of the effectiveness of treatments expressed as single study contrasts as well as summary treatment effects for each outcome. The forest plots presenting the current study findings are based on the results of induvial inverse variance weighted random effects meta-analyses conducted for each outcome.

Effect-size coefficients provide Log Odds Ratio (LOR) information. LORs are weighted effect size coefficients of the treatment impacts for each study corrected for small sample size error. LORs correct for the influence of sample size differences and weight the estimates of single studies relative to other studies included in the sample. Similar to ORs, LORs are interpreted as likelihood that a specific outcome will or will not happen dependent on participation in a treatment group or comparison group (i.e., control group). In the current study, treatment groups are exposed to restorative interventions, whereas control groups are exposed to traditional criminal justice system processes. Using organization and symbology, forest plots graph the distribution of LORs in terms of size, statistical significance, and consistency (i.e., confidence interval overlap).

Forest plots are constructed using two perpendicular lines. The horizontal lines with boxes represent the outcome measure (e.g., log odds of recidivating, log odds of being satisfied with intervention, log odds of complying with restitution). Whereas the vertical line (i.e., line of no difference) intersects the horizontal line and represents where the intervention had no effect. The LOR estimates of studies which appear on the vertical line represent interventions that had no difference on the outcome measure between treatment and control groups. For LORs the line of no difference stands at a value of 0, which is equivalent to no effect, or no association between a treatment and an outcome. The size (i.e., magnitude) of an effect is depicted as a square for each study and the distance of each square from the line of no difference corresponds to the statistical significance of the observed effect. The standard error of each effect-size coefficient, 95% confidence interval (CI), are represented as horizontal lines that perpendicularly run through their corresponding estimates. The pooled effect estimate takes the form of a diamond at the bottom of the plot. The peaks of the summary effect diamond fall at the point estimate value

and the horizontal points represent the upper and lower confidence intervals of that effect. Study identification and tabled information, such as LOR point estimate, confidence intervals, z values, and p values, are also provided for interpretation of results.

Recidivism. A total of 31 effect-sizes were pooled to calculate a summary effect of restorative justice programs on subsequent arrest outcomes. Figure 3 displays a forest plot of LORs comparing effects between treatment and control groups for each study. Effect sizes which fall below 0 indicate a negative effect for recidivism (i.e., less subsequent arrests). The distribution of LORs in Figure 3 indicates mixed treatment effects, with some individual findings in favor of the treatment intervention and some results in favor of control group participation (i.e., treatment as usual). However, the summary treatment effect based on the combined populations across studies in the sample provides moderate support for the effectiveness of restorative justice programs.

Study name	Log odds ratio	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Brooks, 2013	1.027	0.206	0.043	0.623	1.431	4.980	0.000
Cosden et al., 1999	0.538	0.427	0.183	-0.300	1.375	1.259	0.208
Povitsky Stickle et al., 2008	0.325	0.430	0.185	-0.518	1.169	0.756	0.449
Crofoot, 1987	0.314	0.281	0.079	-0.237	0.864	1.117	0.264
Lincoln et al., 1977	0.123	0.364	0.133	-0.591	0.837	0.338	0.736
Jeong et al., 2012	-0.029	0.173	0.030	-0.368	0.309	-0.169	0.866
Davis, 2009	-0.030	0.459	0.211	-0.930	0.871	-0.064	0.949
Severy & Whitaker, 1982	-0.035	0.146	0.021	-0.321	0.251	-0.239	0.811
Gase et al., 2016	-0.114	0.306	0.094	-0.713	0.486	-0.372	0.710
Shelden, 1999	-0.213	0.197	0.039	-0.600	0.173	-1.081	0.280
Mills, Barocas, & Ariel, 2012	-0.281	0.414	0.171	-1.092	0.529	-0.680	0.497
Hitao, 1999	-0.305	0.615	0.378	-1.510	0.900	-0.497	0.619
Griffith, 1983	-0.314	0.242	0.058	-0.788	0.159	-1.301	0.193
Lee, 1999	-0.360	0.279	0.078	-0.907	0.188	-1.288	0.198
McGarrell et al., 2000	-0.456	0.260	0.068	-0.966	0.054	-1.754	0.079
Bergseth & Bouffard, 2007	-0.577	0.369	0.136	-1.301	0.147	-1.563	0.118
Quay & Love, 1997	-0.578	0.203	0.041	-0.977	-0.180	-2.844	0.004
de Beus & Rodriguez, 2007	-0.608	0.049	0.002	-0.704	-0.513	-12.474	0.000
Umbreit, 1994a	-0.621	0.364	0.133	-1.335	0.092	-1.706	0.088
Wiinamaki, 1997	-0.664	0.225	0.051	-1.106	-0.223	-2.952	0.003
Patrick et al., 2004	-0.751	0.710	0.504	-2.142	0.641	-1.058	0.290
Henggeler et al., 1992	-0.775	0.445	0.198	-1.647	0.097	-1.741	0.082
Berger et al., 1977	-0.782	0.105	0.011	-0.987	-0.577	-7.466	0.000
Duwe, 2018	-0.887	0.481	0.232	-1.831	0.056	-1.843	0.065
Shichor & Binder, 1982	-0.906	0.632	0.399	-2.145	0.332	-1.434	0.152
Carr & Nelson, 2000	-0.926	0.330	0.109	-1.573	-0.278	-2.803	0.005
Fors & Rojek, 1999	-1.034	0.250	0.062	-1.524	-0.544	-4.138	0.000
McGarrell & Kroovand Hipple, 2007	-1.067	0.167	0.028	-1.395	-0.739	-6.375	0.000
Myers et al., 2000	-1.186	0.735	0.541	-2.627	0.256	-1.612	0.107
Quinn & Van Dyke, 2004	-1.225	0.212	0.045	-1.641	-0.809	-5.777	0.000
Dick, 1999	-1.414	0.589	0.347	-2.569	-0.260	-2.401	0.016
	-0.415	0.094	0.009	-0.599	-0.230	-4.412	0.000

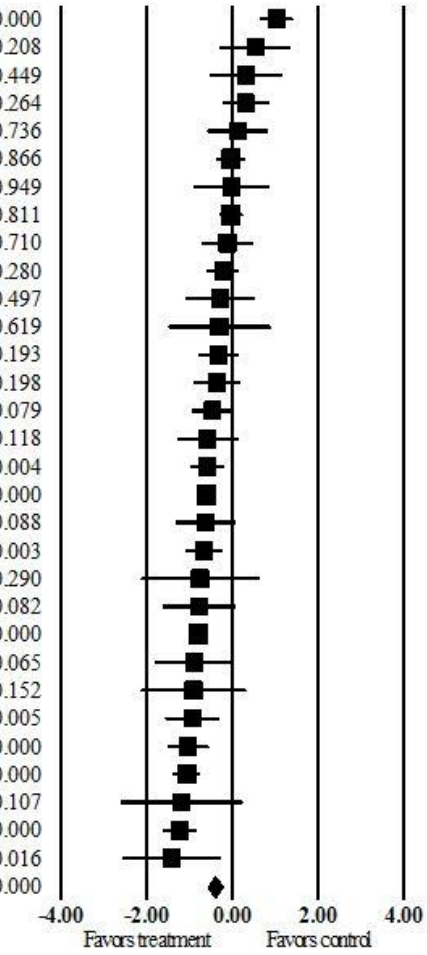

Figure 3. Meta-analysis of subsequent arrest LOR outcomes using random effects models

The LORs for arrest outcomes for treatment offenders ranged from -1.414 (1.41 times less likely to be arrested) to 1.027 (1.02 times more likely to be arrested). Treatment effects for 5 of the 31 studies (16.01%) yielded greater recidivism among treatment groups but only 1 of these effects was statistically significant ($p < 0.000$). Of the 26 studies reporting less recidivism among treatment groups, 9 were statistically significant ($n = 5$, $p < 0.05$; $n = 4$, $p<0.001$). The net impact of restorative justice programs on the odds of subsequent arrest were in favor of restorative interventions. The summary effect estimate supports a moderate impact on subsequent arrest outcomes with an LOR effect size of -0.415, 95% CI [-0.599, -0.230], $z = -4.412$, $p < 0.000$ (see diamond at the bottom of Figure 3). The estimated Q statistic of 146.320 ($df = 30$, $p < 0.000$) suggests that the 31 effect-sizes are highly heterogeneous. Similarly, the I^2 value revealed that 79.50% of the variance among sample effect-sizes was not random. In other words, the results of the meta-analysis of recidivism outcomes indicates that restorative justice programs have statistically significant lower subsequent arrest rates than their control counterparts. Offenders who were exposed to the restorative justice treatment were 41.5% less likely to be arrested post intervention when compared to offenders processed through the traditional criminal justice system. However, substantial heterogeneity found within the sample must be addressed to more holistically understand these findings. The author was also interested in determining how satisfied victims were with their experiences of restorative justice compared to victims within the criminal justice system.

Victim Satisfaction. Victim Satisfaction was examined in 37 (36.7%) of the original 121 study sample for dependent variable coding. Of these 37 articles, the

majority (83.7%) failed to measure victim satisfaction outside of treatment group participation, uses matching techniques to create a comparison group, report data needed for effect-size calculation, or simply did not report quantitatively on this outcome. Due to these limitations, only six studies were included for meta-analysis. Only 2 studies are needed to conduct a meta-analysis, however this sample was considered underpowered, particularly for assessing heterogeneity (Valentine et al. 2010). Therefore, the results of this meta-analysis should be taken with caution. Unlike the interpretation of subsequent arrest outcomes, effect sizes for victim satisfaction (i.e., LORs) above zero would be expected in favor of treatment. Figure 4 provides meta-analytic results for victim satisfaction outcomes among victims exposed to restorative interventions as compared to criminal justice system treatment as usual.

Study name	Log odds ratio	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Dick, 1999	2.909	1.274	1.624	0.411	5.407	2.282	0.022
Hitao, 1999	2.398	1.727	2.982	-0.987	5.782	1.389	0.165
Carr & Nelson, 2000	1.094	0.311	0.097	0.485	1.703	3.519	0.000
Umbreit, 1994a	1.014	0.218	0.047	0.587	1.441	4.653	0.000
Cosden, Casas, & Wolfe, 1999	0.405	1.683	2.833	-2.894	3.705	0.241	0.810
Lane, Turner, Fain, & Sehgal, 2005	-0.263	0.230	0.053	-0.713	0.188	-1.144	0.253
	0.867	0.418	0.175	0.047	1.687	2.071	0.038

Statistics for each study

Log odds ratio and 95% CI

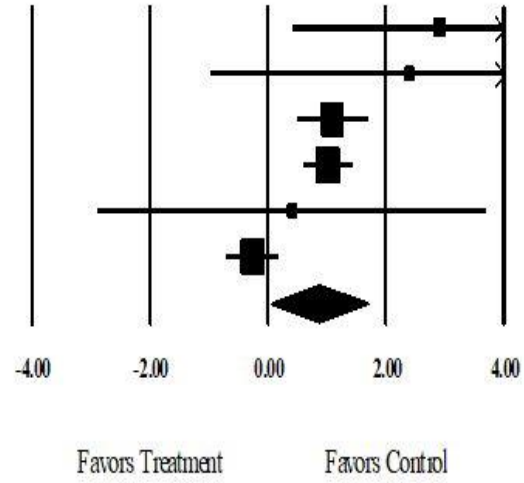

Figure 4. Meta-analysis of victim satisfaction outcomes using random effects models

The odds of victims being satisfied with their participation in the treatment were overwhelmingly positive (LOR = 0.867, 95% CI [-0.263, 2.909], z = 4.165, p < 0.000 (see diamond at the bottom of Figure 4). The LORs for victim satisfaction ranged from -0.263 to 2.909. Results favored the intervention in terms of victim satisfaction with their restorative justice program participation. Five out of six studies yielded positive effects. Three out of these five were statistically significant at the $p < 0.05$ level and 2 were statistically significant at the $p < 0.001$ level. The treatment victims in four studies reported being satisfaction at odds as much as one to two-and-a-half times greater than their counterparts. These findings mean that in the overall sample, treatment victims were 86.7% more likely to feel satisfied with the restorative justice process when compared to victims exposed to traditional criminal justice practices. Heterogeneity was low among studies in the sample (Q = 7.678, df = 5, p = 0.175, I^2 = 34.88%). Therefore, the chance that random factors influenced these findings is minimal.

Restitution Compliance. Restitution compliance was the third outcome of interest in this meta-analytic investigation. Six studies were used to compare the completion of restitution by offenders between restorative treatment groups and control groups. Conclusions drawn from the meta-analysis findings of this restricted sample were underpowered, which limits the research's ability to generalize results. Fulfillment of restitution was measured dichotomously such that the number of offenders who completed their restitution obligations were reported as events in each group. Instances of offenders complying with restitution were summated for a total of compliances per treatment and control groups for each study. This information was used to generate individual treatment contrasts and a summary effect coefficient (LORs). More compliance in intervention groups was found in five out of six studies in the sample. Figure 5 presents the results of the meta-analysis for 6 studies on the outcome of restitution compliance.

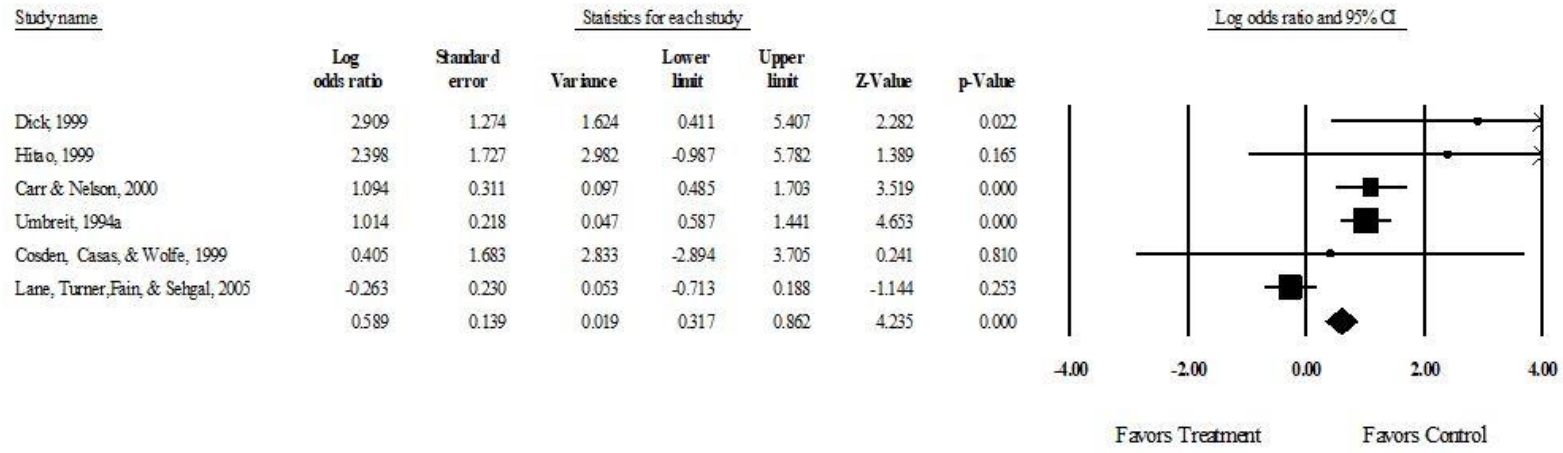

Figure 5. Meta-analysis of restitution compliance outcomes using random effects models.

Effect sizes (LORs) ranged from -0.263 to 2.909. Results favored restorative interventions regarding offender compliance to restitution obligations. Five out of six studies reported positive restitution outcomes, 1 of which was significant at the $p < 0.05$ level and 2 of which were statistically significant at the $p < 0.001$ level. One study observed greater instances of restitution compliance for control participants, but difference in LORs were not statistically significant ($p = 0.235$). Overall, the odds of treatment group offenders completing their restitution obligations to victims were great (LOR = 0.589, 95% CI [0.317, 0.862], $z = 4.235$, $p < 0.000$) (see diamond at the bottom of Figure 5). Four out of the 6 studies yielded LORs which correspond to increased odds of compliance to restitution as much a 1 to 2.9 times more likely to fulfill restitution in treatments compared to control groups. Overall, treatment offenders had 58.9% greater odds of completing their restitution than offenders not exposed to the restorative interventions. Conservative consideration should be given to the summary effect due to the high levels of heterogeneity in the sample. Testing revealed that an estimated 79.68% of the variation in treatment effects was not of common cause ($Q = 24.601$, $df = 5$, $p < 0.000$). However, these results may be positively skewed by studies which reported 100% offender compliance with restitution agreements in their treatment groups ($n = 2$).

Desired effects of participation in restorative justice intervention were evident across all three success outcomes in the current sample. Comparing treatment participants to control participants, offenders had great odds of avoiding subsequent arrest and fulfilling restitution to victims, and victims were more likely to feel satisfied by the restorative process. However, varying levels of heterogeneity found within two of three samples was cause for concern. Moderating tests were needed to examine the likelihood that observed effects were due special

cause variation (i.e., not due to chance). Furthermore, underpowered samples in two of the three meta-analyses require greater scrutiny for publication bias due to small sample size.

Subgroup Analysis

The notable heterogeneity found among studies used for meta-analyses warranted further examination to detect for potential moderating influences. However, due to the limited sample-size of victim satisfaction and restitution compliance evaluations, only heterogeneity testing of subsequent arrest results was feasible in the current study. To assess potential moderating effects for these results, four categorical indicator variables were generated to capture different study characteristics. The creation of each subgroup was guided using suggestions for further research of extant literature as well as original inquiries of the current investigation.

Prior research has suggested that study characteristics such as program type (e.g., mediation, restitution, etc.), offender voluntary participation (i.e., completely voluntary, somewhat voluntary, mandated), offense type (e.g., property, violent, etc.), and offender type by age (i.e., juvenile offenders, adult offenders) may moderate the success of restorative justice programs (Coates & Brown, 2003; Crofoot, 1988; Mika et al., 2004; Van Wormer, 2009). Despite scholars maintaining the importance of considering these factors, little research has directly examined their moderating relationships (Bouffard, Cooper, & Bergseth, 2017; Strode, 1997). Unique to the current study, subgroup analysis was also used to compared the SMS scores of studies and examine potential relationships between methodological quality and program success (i.e., LORs of subsequent arrest). Studies with the strongest designs (i.e., final SMS score categories 3, 4, and 5) were broken into 3 subgroups to determine whether LORs significantly differed based on objective study rigor. Table 12 presents the LOR results of subgroup analyses relative to independent variable groupings on subsequent arrest outcomes. All

indicator variables but one (i.e., offense type) significantly moderated the effects of restorative interventions on subsequent arrest (see Table 12).

Table 12.

Subgroup analysis of study characteristics on subsequent arrest outcomes

Indicator Variables	LOR, z-value, p-value
*__Type of Program__ (n= 31)	
Circles (n = 2)	LOR = -0.538, z = -1.717, p = 0.086
Mediation (n = 8)	LOR = -0.186, z = -0.602, p = 0.547
Conferencing (n = 6)	LOR = -0.588, z = -2.387, p = 0.017
Restitution (n = 4)	LOR = -0.330, z = -1.397, p = 0.162
Hybrid (n = 5)	LOR = -0.683, z = -4.923, p = 0.000
Other (n = 6)	LOR = -0.367, z = -1.958, p = 0.050
*__SMS Score__ (n = 31)	
3 (n = 9)	LOR = -0.507, z = -4.767, p = 0.000
4 (n = 14)	LOR = -0.782, z = -1.583, p = 0.113
5 (n = 7)	LOR = -0.745, z = -1.839, p = 0.000
__Offense Type__ (n = 15)[a]	
Non-violent, property (n = 4)	LOR = -0.468, z = -4.389, p = 0.000
Violent (n = 1)	LOR = -0.755, z = -0.680, p = 0.497
Sex crime (n = 1)	LOR = -0.412, z = -1.843, p = 0.065
Mixed (n = 16)	LOR = -0.649, z = -3.399, p = 0.001
*__Offender Age__ (n = 31)	
Juvenile (n = 25)	LOR = -0.637, z = -4.511, p = 0.000
Adult (n = 6)	LOR = -0.876, z = -0.309, p = 0.756
Overall	LOR = -0.648, z = -4.463, p = 0.000
*__Offender Participation__ (n = 31)	
Voluntary (n = 9)	LOR = -0.248, z = -0.904, p = 0.366
Somewhat voluntary (n = 13)	LOR = -0.365, z = -3.203, p = 0.001
Mandated (n = 9)	LOR = -0.689, z = -5.323, p = 0.000

*Q statistic is statistically significant at the $p < 0.001$ level.

a. 7 studies did not specify offense types.

Restorative justice programs are predominately used with juvenile offenders (Roach, 2000; Umbreit, 1994; Zejr, 2015). This trend was reflected in the current study sample. Twenty-five out of 31 programs evaluated in the meta-analysis sample exclusively included juvenile offender participants. Although the variability in the sample by offender type was not large, differences in results between offender type significantly moderated the impact of restorative justice programs on subsequent arrest of offenders ($p = 0.000$). Juveniles had a greater likelihood of avoiding arrest post intervention than adult offenders (LOR = -0.288, $p = 0.000$ and -0.450, $p = 0.422$ respectively). This means the odds of having subsequent arrests were 17% greater for adults than for juvenile offenders. However, this finding is limited due to the sizeable difference in number juvenile included in the meta-analysis ($n = 25$) compared to the proportion of adults in the sample ($n = 6$). Therefore, no definite conclusion can be drawn about how well restorative justice serves one population over the other in terms of avoiding re-arrest.

Similarly, the quality of research design also significantly moderated the success of restorative justice programs in the United States ($p = 0.000$). Methodological strength groupings accounted for a sizable percent differential (about 27.5%) in the decreased odds of subsequent arrests. As anticipated by the author, methodological strength accounted for some between-studies heterogeneity ($p < 0.001$) such that the poorest methodology (SMS = 3) was associated with substantially greater odds of arrests (LOR = -0.507, $p = 0.000$) followed by studies with the most rigors methods (SMS = 5, LOR = -0.745, $p = 0.000$). The middle methodological strength category of 4 achieved the most favorable odd for avoiding arrest with treatment offenders 78.2% less likely to be subsequently arrested than control participants, but this difference was not statistically significant ($p = 113$). In other words, studies with an SMS score of 5, corresponding the highest methodological quality, had treatment offenders who were 74.5% less

likely to recidivate compared to their control counterparts. These results illustrate the importance of sound methodological testing of restorative justice impacts on determining their utility as an alternative to traditional responses to crime. Researchers should consider the objective methodological strength of the studies that they include in meta-analysis to avoid misleading results based on flawed methodology. Any findings derived from a sample utilizing lesser quality studies should be taken with caution.

Significant differences were observed between outcomes of studies evaluating different restorative justice programs ($p = 0.001$). Of the traditional program types (i.e., mediation, conferencing, circles, and restitution), mediation seemed to have the least favorable impact on subsequent arrest outcomes ($n = 8$, LOR=-0.187, $p =0.547$), whereas conferencing had the greatest impact ($n = 6$, LOR = -0.588, $p = 0.017$). The remaining two traditional treatment modalities of circles and restitution also yielded positive subsequent arrest outcomes with offenders in treatment groups less likely to reoffend than control groups by 8.6% and 16.2% respectively. However, the latter findings are limited due to the small prevalence of circle programs in the sample and lack of statistical significance ($n = 2$; $p = 0.086$ and $n = 4$, $p = 0.162$ respectively). These results are further contextualized when the moderating effects of emergent program categories are considered.

Hybrid programs, those restorative programs which include more than one treatment modality, displayed the best recidivism outcomes of all programs included in the meta-analysis sample ($n = 5$, LOR = -0.683, $p = 0.000$). Hybrid offenders were 68.3% less likely to be arrested post intervention compared to control group offenders. Similarly, the other programs category, which capture interventions models with elements of restorative justice as well as non-traditional restorative justice programs, resulted in fewer arrests in treatment compared to control groups.

Individuals exposed to other category interventions were 36.7% less likely to be arrested after participation than their control counterparts (n = 6, LOR = -0.367, p = 0.05). Taken together the results of subgroup analysis examining the influence of program type on recidivism detected significant moderating effects based on program type. With exception to the category of circle programs, the distribution of program modalities was fairly even, meaning that these findings can be taken with some confidence.

Finally, the author predicted that voluntary offender participation would significantly account for variation within treatment outcomes. While past research has demonstrated the negative effects of selection bias on program success, poor outcomes have also been observed for offenders mandated to participate in restorative justice programs. In the current study voluntary participation did significantly moderate the effectiveness of restorative justice on subsequent offender arrest (p = 0.000). Dissimilar to prior research, mandated offender participation was associated with the greatest reduction to future arrest. Offenders who ordered to participate (n = 9) in restorative programs were 68.9% less likely to be arrested following intervention than control group participants (p = 0.000). Somewhat voluntary participation, differentiated from the "voluntary" category by offender declining to participate or failing to complete programing resulting in a return the criminal justice system and/or legal consequences, also yielded positive effects for subsequent arrest (LOR = -0.365, p = 0.001), followed by completely voluntary participation which was favorable but not statistically significant (LOR= -0.248, p = 0.366). In the offender participation grouping, reductions in the odds of arrest for treatment offenders ranged from being a about a quarter to two-thirds less likely to be subsequently rearrested than control offenders. However, these findings may be limited in their generalizability due to a skewed distribution in voluntary participation categories.

Other limitations within this grouping constrain the researcher's abilities to infer difference between participation groups. Although the greatest outcomes were associated with mandated participation and the poorest outcomes were associated with entirely voluntary participations, both groups made-up the minority of studies (9 studies each). Furthermore, the successful subsequent arrest outcomes found for the somewhat voluntary participation group are complicated to interpret. Somewhat voluntary participation may include offenders who would otherwise have chosen to participate (similar to those who chose to participate in the completely voluntary category) or may have included offenders who would have otherwise declined to participate if doing so did not result in a return to traditional criminal justice processing. Therefore, how important voluntary participation is to the effectiveness of restorative justice programs in the United States remains an open question.

Publication Bias Testing

Funnel plots were generated to visually review existent publication bias in the meta-analysis samples. Samples devoid of publication bias feature study plots which fall within the pyramid-like shape depicted within a funnel plot. Unlike forest plots which organize effect-size estimates horizontally, funnel plots organize the standard error of effect size data vertically (i.e. SE of LORs). Higher-powered, more precise studies with larger samples are placed at the top and LORs of smaller studies are placed at the bottom. This visual representation helps to identify asymmetry due to studies which may be missing from the sample as an effect of selective publication (i.e., file drawer problem), or other faulty sampling issues.

Figure 6 presents a funnel plot of LOR effect estimates for the outcome measure of subsequent arrest. While it is clear from this graphic that there may be evidence of variation not due to chance (i.e., special cause variation), the general shape is consistent with a pyramid. The

shape and distribution of the plots coupled with the results of the moderator testing suggest some between study heterogeneity. However, no evidence of publication bias was found using either Egger's regression intercept ($t = 1.105$, $p = 0.139$) or Begg's test (adjusted Kendall's Tau = -15, $z = 0.255$, $p = 0.399$). Duval and Tweedie's trim and fill method specified adjusted values for 7 studies missing from the right of the mean (LOR = 0) (see Figure 7).

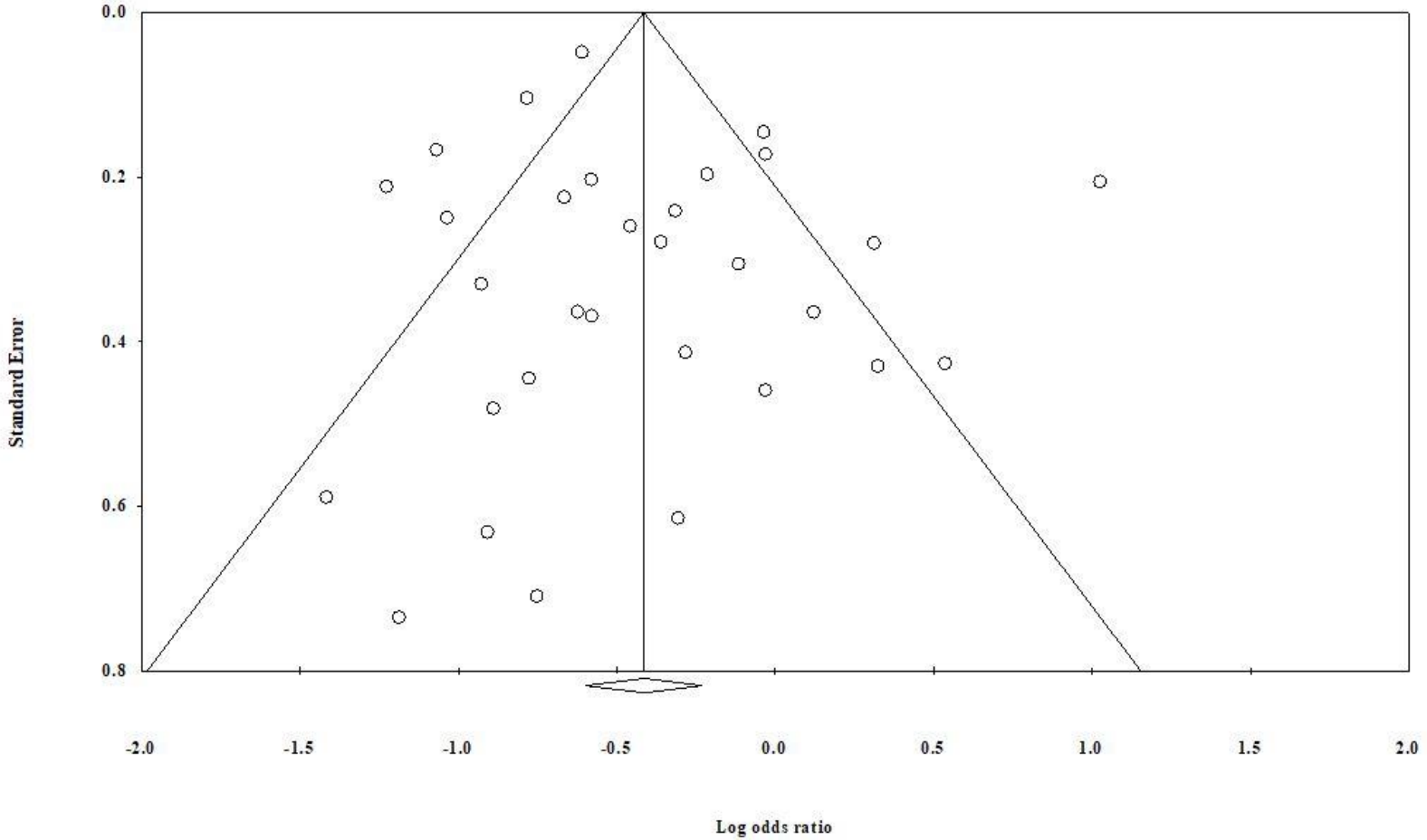

Figure 6. Funnel plot of standard rrror by LORs of subsequent arrest studies

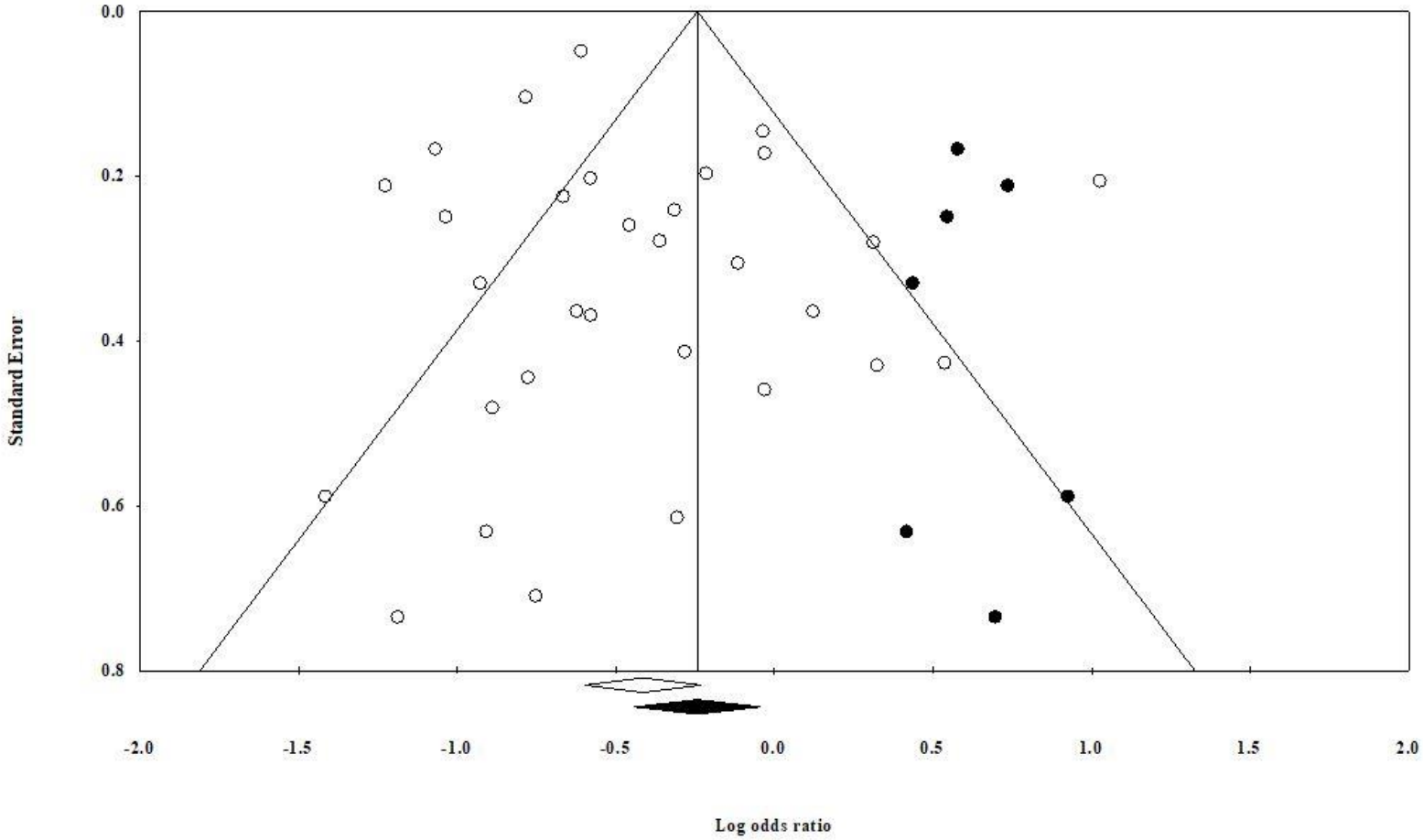

Figure 7. Trim and fill funnel plot of standard error with observed and imputed LORs of subsequent arrest studies

Figure 8 displays the standard error for victim satisfaction LORs. One outlier is apparent in the funnel plot. However, results for Kendall's Tau with continuity correction (adjusted Kendall's Tau = -3, z = 0.564, p = 0.287) and Egger's Test (t = 0.775, p = 0.241) indicate that the sample did not suffer from publication bias. The publication bias tests for restitution compliance were also favorable. Trim and fill results are plotted on Figure 9. Only one study was specified as missing from the sample. Figure 10 presents a funnel plot of standard error by LOR with one study suggesting special cause variation. Similar to victim satisfaction results, publication bias testing did not indicate the presence of missing articles within the restitution compliance sample. Both Kendall's Tau with continuity correction (adjusted Kendall's Tau = -3, z = 0.376, p = 0.354). or Egger's Test (t = 0.775, p = 0.241) were insignificant. Figure 11 displays the trim and fill results for restitution compliance studies with two imputed studies on the bottom left of the plot. The results of publication bias tests for both victim satisfaction and restitution compliance samples should be taken with caution due to their sensitivity to small sample size. Therefore, these tests may have been under-powered to detect bias.

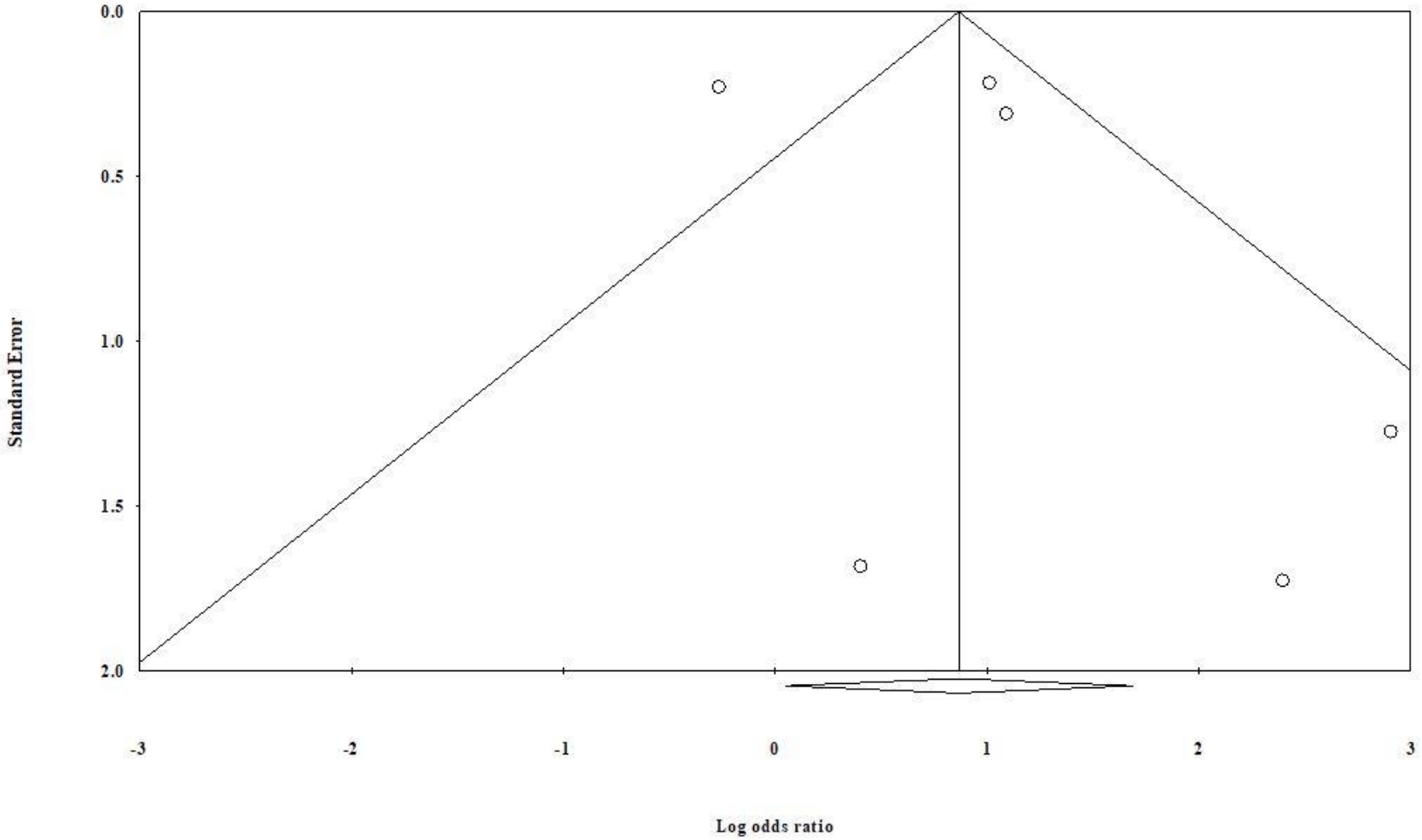

Figure 8. Funnel plot of standard error by LORs of victim satisfaction studies

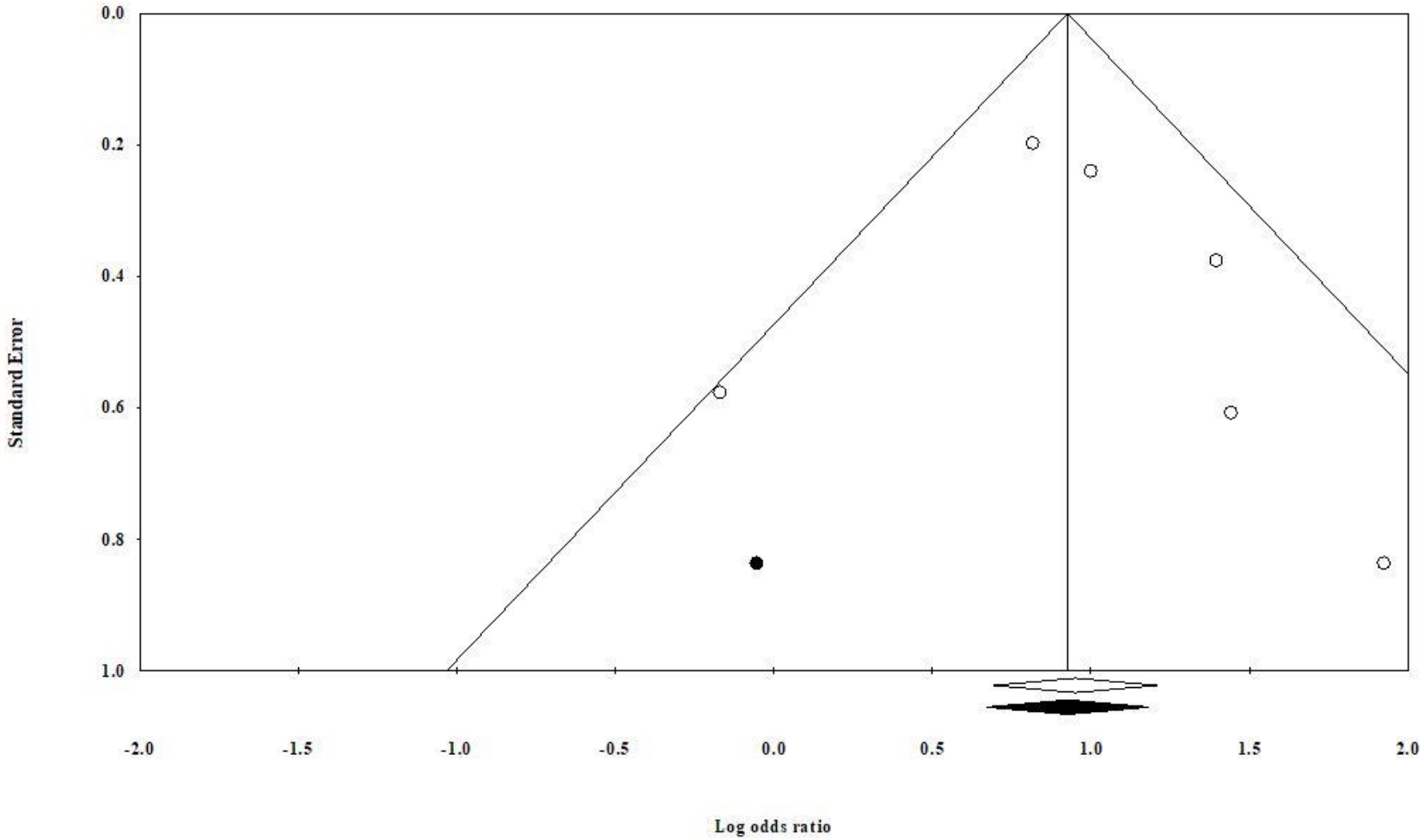

Figure 9. Trim and fill funnel plot of standard error with observed and imputed LORs of victim satisfaction studies

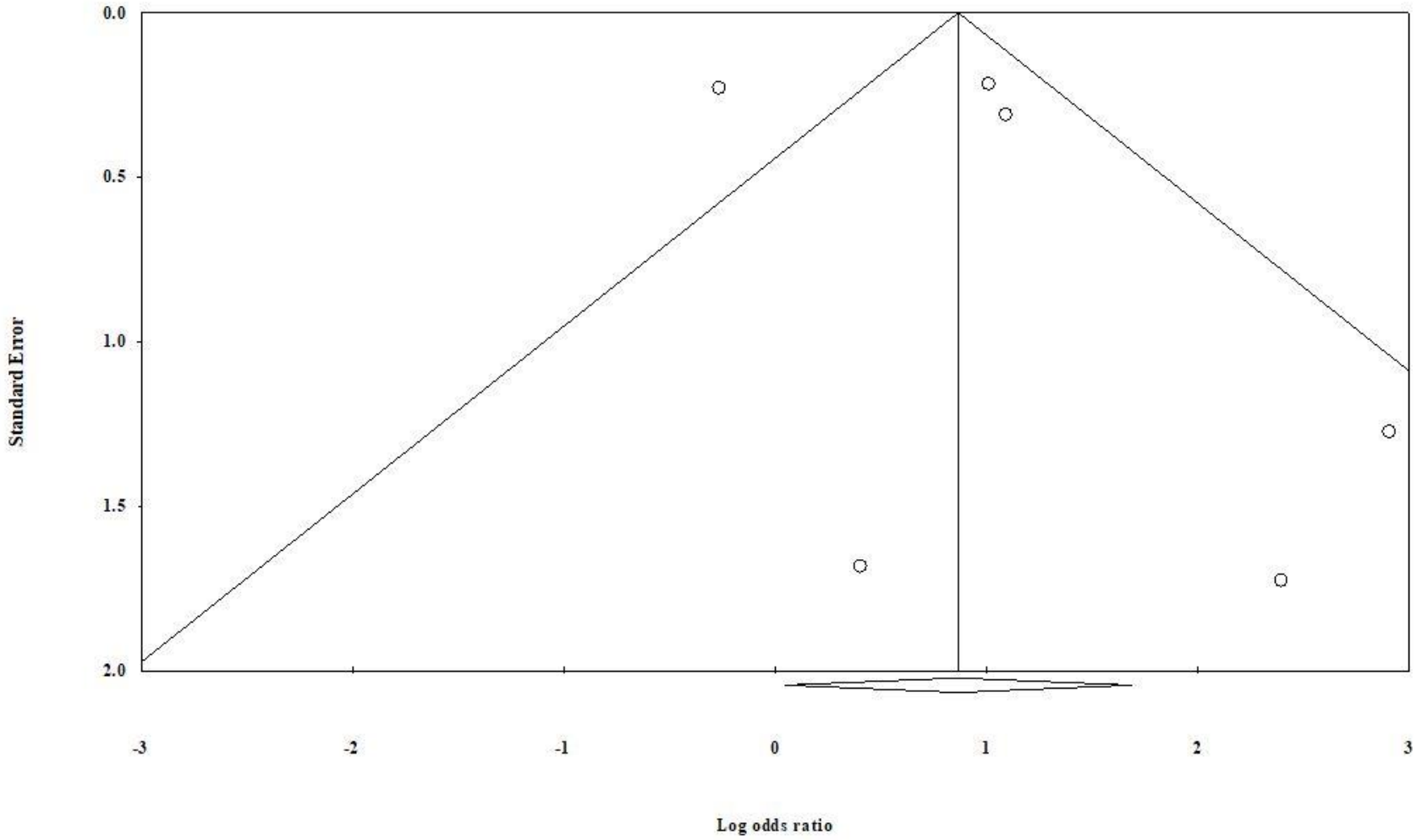

Figure 10. Funnel plot of standard error by LORs of restitution compliance studies

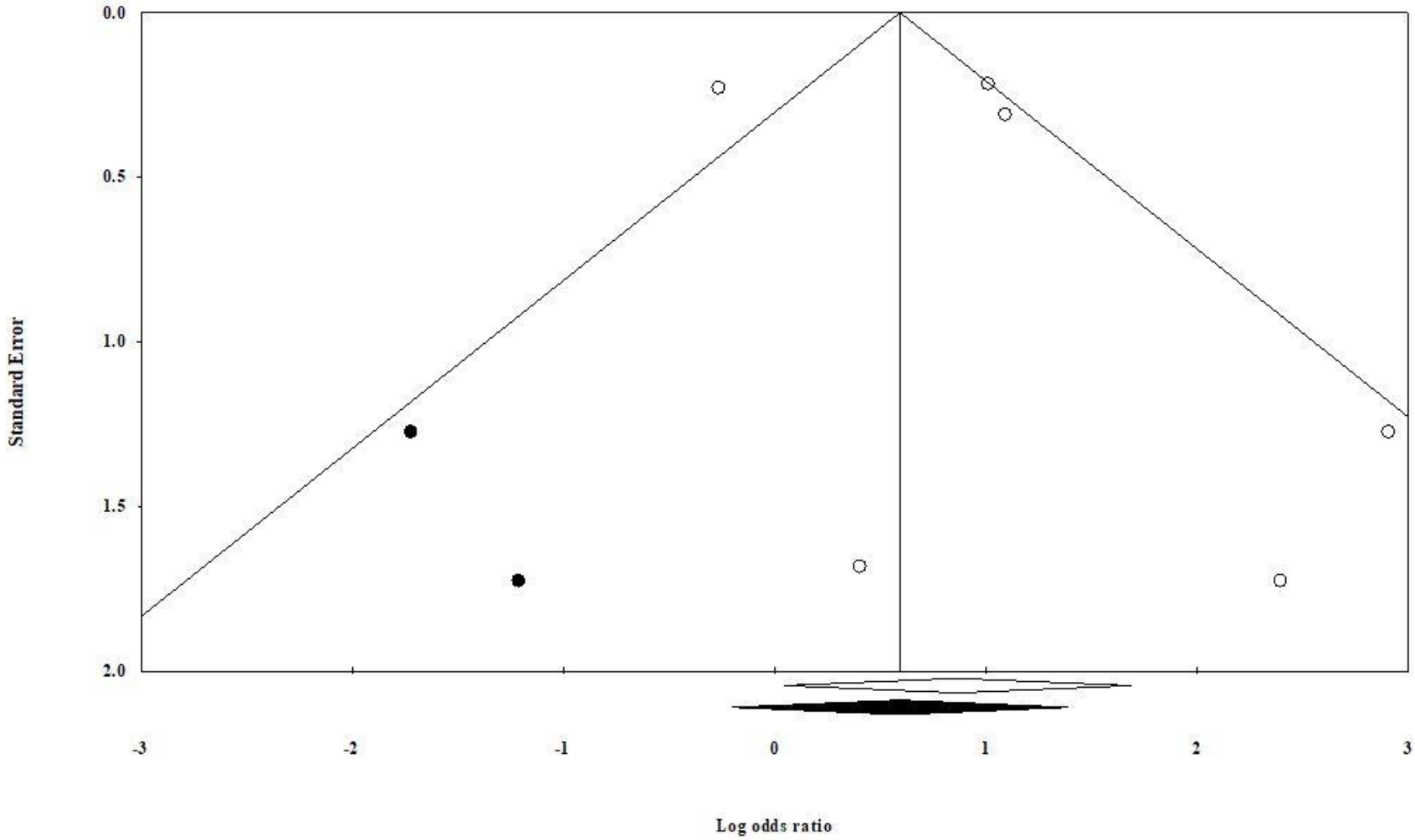

Figure 11. Trim and fill funnel plot of standard error with observed and imputed LORs of restitution compliance studies

Discussion

Efforts to evaluate restorative justice programs as well as the sophistication of evaluative research designs have intensified with its increased prevalence in the United States since the 1970s (Barnett, 1977; Beale, 2003; Kurki, 2000; Latimer et al., 2005; Pavelka, 2016). Yet, extant empirical research provides mixed support in favor of restorative justice applications and has not considered the objective quality of past research methodology (Bergseth, & Bouffard, 2007; Bonta et al., 2006; Nugent et al., 2003, 2004; Poulson, 2003). The scope of this research sought to provide clarity about the success of restorative justice within the United States as well as critically examine the foundation of literature upon which claims of programmatic effectiveness are made. This investigation examined measures of success and the rigor of research designs as well as assessed the effectiveness of restorative justice programs to decrease subsequent arrests, increase victim satisfaction, and increase restitution compliance.

Research Aim One

Prior evaluations have assessed restorative justice program success using diverse measurements (Hotaling & Buzawa, 2003; Parsons & Bergin, 2010; Presser & Van Voorhis, 2002; Umbreit,1989). A systematic search protocol identified 121 evaluations for review of dependent measures of restorative justice success. Across these studies, more than 60 different operational definitions were employed. The author's hypothesis, consistent with prior literature, was the reliance of recidivism as a proxy for program success (i.e., subsequent arrest). Recidivistic outcomes were measured most often in studies (97.5%), followed by satisfaction outcomes (66.1%), and restitution compliance outcomes (36.7) (see Table 9). However, recidivism as a measure of success is flawed

due to lack of standardization of measurement as well as its accuracy to capture offender rehabilitation (Blumstein & Larson, 1971; Harris et al., 2011; Mair, 1991; Maltz, 1984).

Program evaluations' emphasis on recidivism is likely influenced by traditional ideas of what offender rehabilitation encompasses (i.e., desistence from criminality as behavioral change). However, past research documents the limitations of using recidivism as a measure of effectiveness (Mair, 1991; Maltz, 1984; Presser & Van Voorhis, 2002). Although the official records of re-offending are readily available and intuitively appealing compared to harder to gather information related to the measurement of mechanisms and other examples of offender change (e.g., empathy, guilt, self-esteem, self-identity), researchers may be overlooking important impacts for the offender which cannot be assessed by recidivism alone.

In regard to victim impacts, several evaluators attempted to assess victim recovery but did so in vastly different ways. For example, victim recovery measures examined aspects as diverse as fear of revictimization, changes in victim offender relationships, and mental health outcomes. Because of the multifaceted nature of victim recovery, standardization of measurement may not be possible or advisable. Instead, research might start by first asking victims what is meaningful to their recovery and then attempt to measure how well restorative justice meets those specified needs. This observation is in line with prior research which notes the value of qualitative and mixed-methods data collection within restorative justice impact evaluations (Presser & Van Voorhis, 2002).

Less than 5% of the studies measured success related to stakeholder relations/ socialization, emotions, and perceptions, not including victim satisfaction. This oversight is particularly concerning considering that the goals of restorative justice are focused on

redressing harms and broken relationships. Although some evaluators modeled their measurement of success using the express goals of restorative justice (e.g., perceived repair of harms, improved relationships) (n = 18), other studies examined the byproducts of restorative justice goals (e.g., issuing of an apology, restorative contract created) (n = 12). These outcomes are more in-line with past literature exploring alternative measures of offender rehabilitation (Choi & Severson, 2009; Tangney et al., 2011). Finally, numerous studies measured success in terms consistent with program components (e.g., perceived engagement, involvement, opportunity to express views) (n = 16).

From examining the outcome measures of 121 studies, it is clear that the victims and community members who participate in the restorative process are generally not prioritized ahead of offenders. Aside from minimal experiential outcome measures such as support stakeholder and community stakeholder satisfaction (n = 13), support stakeholders (i.e., parents, peers, citizens) were excluded from the evaluation process entirely. By comparison about three-quarters of studies examined at least one outcome focused exclusively on victims or offenders (n= 89, 73.6%). While some programs may not have included community members in the process because they were not relevant, such as with restitution and certain other forms of restorative justice (e.g., victim impact panels), the majority of evaluations were conducted using dialogue-driven interventions (i.e., mediation, circles, conferencing) which encourage the inclusion of citizens as active participants. Not a single study examined community-level impacts of restorative justice. Without measurement of community impact, it is impossible to draw conclusions about true restorative justice program effectiveness. Success measures need to be more inclusive of various stakeholders so as not to overlook their experiences and outcomes

but must also work to move beyond micro-level indicators of success to capture community-level outcomes. When taken together, these results indicated that community-oriented outcomes are either too methodologically difficult to capture or are not meaningful for how success is currently being conceptualized by researchers.

Research Aim Two

The second main finding of this study concerned the methodological strength among the sample of 121 impact evaluations. The breakdown of SMS scores revealed that the relative methodological rigor of restorative justice evaluations varied considerably. The author predicted there to be fewer studies of high methodological quality. However, scores were generally well distributed across all five score categories. Employing the more restrictive scoring guidelines for SMS was also meaningful for study findings (see Appendix B). The smallest proportion of studies earned the highest methodological strength ranking after score adjustments (9.09%) and the largest number of studies were scored in the lowest SMS category (30.57%). Had raw scores not been adjusted using point deductions for methodological shortcomings the composition of SMS scores within the current study sample would be notably different. Before score adjustments the proportion of SMS scores of 5 was doubled (18.18%), supporting the importance of greater scrutinization of research methodology.

Interpreting methodological design as a proxy for its quality is not enough. More thorough review of potential methodological violations among program evaluations is important, especially when weighing value of individual study results relative to each other. The most common defects in evaluations were related to sampling. For example, success of participant randomization, quality matching of samples, and avoiding or

accounting for attrition were problematic among studies and these limitations required SMS score reductions. One interpretation of the generally low SMS scores supports the need for more randomized control trials of restorative justice programs to determine their effectiveness. Moreover, randomized control trials absent of methodological defects are needed to draw accurate conclusions about how well restorative justice works however success is defined.

Research Aim Three

In an effort to improve upon evaluative research design several rigorous studies have been conducted in past research. These studies provided a synthesis of the literature assessing restorative justice effectiveness using meta-analytic techniques toward more reliable and robust research findings. Recidivism and victim satisfaction were among the outcomes of interest in prior meta-analyses (Bradshaw et al., 2006; Hayes & Daly, 2004; Luke & Lind, 2002; Rodriguez 2005, 2007). In general, the findings of these meta-analyses are mixed (Niemeyer & Shichor, 1996; Nugent et al., 2003, 2004; Roy, 1993). However, restorative justice has been found to reduce the likelihood of recidivism when compared to the outcomes of non-restorative control groups by as much as 30% (Bonta et al., 1998, 2002; Bradshaw & Roseborough, 2005; Latiemer et al., 2001, 2005; Wilson et al., 2017). Despite increased empirical attention directed at examining program success, the results of extant meta-analyses are limited in their ability to inform debates about the effectiveness of restorative justice within the United States and have overlooked potential moderating effects (e.g., voluntary offender participation, methodological quality).

The final research objective of the current study was two-fold. First, the current study built upon prior meta-analyses by measuring recidivism, victim satisfaction, and restitution compliance outcomes for restorative justice programs in the United States. Second, heterogeneity between subsequent arrest evaluations was investigated using subgroup analysis to test for moderating influences with program type, voluntary participation, offense type, offender type, and methodological strength. It was hypothesized that findings across all meta-analyses would provide evidence of the effectiveness of restorative justice programs. The results of meta-analyses supported this hypothesis. Exposure to restorative justice reduced the odds of subsequent arrest by 41.5% compared to control offenders. Similarly, victim satisfaction and restitution compliance outcomes were also in favor of treatment over control. Victims were 86.7% more likely to report satisfaction with restorative justice programs compared to their control counterparts. Offenders were 58.9% more likely to fulfill their restitution agreements when they participated in restorative justice programs compared to offenders processed through the criminal justice system. While the subsequent arrest meta-analysis sample was fairly robust, consisting of 31 studies, the remaining meta-analytic samples were significantly underpowered with fewer than ten studies. Therefore, evidence for the observed impact of restorative justice on recidivism is more reliable than evidence of effectiveness of restorative justice programs on victim satisfaction and restitution compliance. However, substantial heterogeneity in the subsequent arrest meta-analysis sample was accounted for by some moderating variables.

Voluntary Participation. Prior research cautions against the implementation of restorative justice programs in instances of serious crime or with uncooperative,

disengaged offenders. Researchers cautioning against the use of restorative justice assert that mandated offender involvement may be counterproductive, even detrimental to the restorative process (Bibas & Bierschbach, 2004; Van Worner, 2009). Contrary to these assertions, mandated offender participation was associated with significantly fewer odds of subsequent arrest. Results reveal that offenders not given a choice in their participation in restorative justice interventions had the lowest odds of being arrested after treatment; treatment offenders were 68.9% less likely to recidivate then control offenders. This finding was constrained for several reasons. The favorable recidivism outcome for involuntary offender participation, coupled with the smaller impact found for completely voluntary offender participation (LOR =-0.248), contribute to the ongoing debate about the importance of offender choice.

There are three additional caveats to these findings. First, the sample distribution was skewed such that the "somewhat voluntary" category accounted for the greatest proportion of studies (n = 13, 41.9%). Second, the 24.8% reduction in the odds of subsequent arrest observed for offenders who volunteered to participate in restorative justice was not statistically significant (p = 0.336). Third, the inclusion of a third offender participation category (i.e., somewhat voluntary) may have confounded the moderating test results. Somewhat voluntary participation entitles offenders be presented with a choice to participate in restorative justice to avoid formal criminal justice processes. However, failure to agree to participate in or failure to complete restorative justice programs results in an offender returning to the criminal justice system. Meaning that the current study's conceptualization of the levels of voluntary participation may not be mutually exclusive. Instead of offenders falling in this latter category may have elected to

participate regardless of the ultimatum they were presented with or may have only been coerced into participation to avoid formal punishment. Without this additional information it is not possible to fully explore category differences or make inferences about how voluntary offender participation impacts future odds of arrest.

Program Type. The influence of program type on subsequent arrest outcomes was also tested using subgroup analysis. The frequency of program types included in the sample were well distributed. General findings were in favor of restorative circles and conferencing modalities along with programs including two or more models of restorative justice (i.e., hybrid programs), which yielded fewer instances of subsequent arrest among treatment offenders. However, the variation accounted for by circle programs was not statistically significant ($p = 0.086$). By comparison the program types associated with the smallest reductions to subsequent arrest were restitution and mediation. Several factors may attribute to poorer arrest outcomes for mediation compared to all other program types.

By design, the goals of mediation are in line with the core objectives of restorative justice: reparation of harms, offender accountability, and restoration of relationships. The nature of these variables does not directly relate to avoiding arrest. Mediation may have also experienced the least success compared to other program types due to program pragmatics which might influence the completion of programs and subsequent outcomes. Second to restorative conferencing, mediation programs require stakeholder preparation prior to intervention. Mediation also requires the collaboration of victims and offenders and equal consideration of all stakeholder needs.

In addition to comparing outcomes between all program types, the author was interested in the potential moderating effects between traditional versus emergent programs. In the current study, traditional restorative justice programs included mediation, circles, conferencing, and restitution. The emergent program categories of hybrid program type and other program type are reflective of a wider adoption of restorative justice practices to fit specific case needs, such as those associated with sexual assault (Kross et al., 2003). Offenders who were exposed to hybrid interventions had 63.8% fewer odds of recidivating than their control counterparts ($p = 0.000$). Similarly, other program participants were significantly less likely to be arrested post treatment than those not participating in restorative justice (LOR = -0.367, 0.05). Hybrid programs may yield the greatest success in terms of offenders avoiding arrest because they are versatile. Similarly, other program types may be best suited for addressing the needs of different types of offenders usually neglected by restorative models most prevalent in the United States (i.e., adult offenders, sex offenders, convicted murderers). Furthermore, programs which implement some restorative elements may result in reduced subsequent offender arrests by combining these elements with other successful interventions, such as psycho-social educational courses on the awareness of harm to victims of a certain offense. However, the practical implications of these results are constrained by the variability of program characteristics and standards of program fidelity.

Research suggests that restorative justice may be a more effective and meaningful response to crime than traditional justice due to its intimate nature and inclusivity (Sullivan & Tifft, 2008; Walker, 2002; Willis & Grace, 2009). The author examined the differences between subsequent arrest outcomes for purer, dialogue-driven treatment

modalities (i.e., mediation, conferencing, circles), and restitution which does not necessitate the inclusion of all stakeholders or direct interaction. Understanding how well these programs, excluding hybrid and other programs, work compared to each other helps substantiate these claims. Although mediation programs yielded the worst subsequent arrest outcomes of all program types, this finding was not statistically significant (LOR = -0.187, p = 0.574). When comparing the remaining programs, results reveal that participation in both conferencing and circles yielded better odds from avoiding recidivism than the odds of subsequent arrest for offenders participating in restitution. However, reduced offs of arrest after treatment for circle programs were not significant.

Offender Type. Restorative justice in the United States is most common among youthful offenders. Restorative interventions are thought to be more beneficial to youth offenders than adult offenders (Hudson, 2002; Shapland et al., 2008). Adults also tend to commit more serious offenses than juvenile offenders, which may be less suitable for restorative approaches (Mika et al., 2004). As prevalence of use with adult offenders and serious crime increases, greater evaluative efforts are needed to examine programmatic impacts for the adult population (Hudson, 2002; Shapland et al., 2008). In the current sample juveniles were about three-quarters (63.7%) less likely to get arrested post-intervention compared to control youth. Whereas adults were 87.6% less likely to experience subsequent arrest compared to their control counterparts. However, the difference in odds of recidivating for adults was not statistically significant (p = 0.765) and juveniles accounted for the majority of the sample (n = 25, 80.64%). Despite these caveats, results indicate that adult subsequent arrest outcomes were no worse than juvenile odds of recidivism.

Offense Type. Dissimilar to moderator program type outcomes, significant differences in LORs were not found between studies including different crime types ($p = 0.325$). Although the overall offense type groups did not significantly account for between study heterogeneity in the sample, two differences were observed for different categories of crime. All four crime categories favored subsequent arrest outcomes for restorative justice programs. Studies examining property crime resulted in the greatest odds reduction for subsequent arrest (LOR = -0.848, $p = 0.000$). Offenders of property crime were more than three-quarters (84.8%) less likely to be arrested than their control counterparts. Whereas violent crime offenders were half (53.8%) as likely to have a subsequent arrest post intervention and studies including a mixture of property and violent offenders observed the poorest arrest outcomes (LOR = -0.532, $p = 0.001$). In other words, unique offense types did not moderate positive subsequent arrest outcomes. To better detect differences in the effects between different crime types future research should explore arrest outcomes for specific crimes.

The differences observed between crime type and subsequent arrest were limited in three ways. First, the sample was skewed, whereby the majority (69.5%) of studies included mixed offenses. Second, no studies were included in the sample with included sexual offenses so no effect could be generated for that crime category. Third, the effect of reduced likelihood of subsequent arrest was constrained by a lack of statistical significance ($p = 0.086$). Based on these findings, conclusions cannot be draw about which crime type is best suited for restorative justice programs. Composition of offense types within this grouping notwithstanding, the success of restorative justice programs was not dependent on the offenses which are included in evaluations. Therefore,

restorative justice should be considered a viable alternative to the traditional criminal justice approach for different times of crime.

Methodological Quality. The author predicted that methodological strength would influence the odds of subsequent arrest such that higher quality methods would be negatively correlated with instances of future arrest among treatment offenders. This prediction was partially supported by the results of subgroup analysis. Objective methodological rigor was found to be a significant moderating variable in the relationship between restorative justice program participation and the likelihood of subsequent arrest. Studies with the weakest methodological strength category (SMS = 3) scores rendered the greatest reduction in the odds of arrest, whereas studies with the strongest methodological designs (SMS = 5) resulted in the poorest subsequent arrest outcomes. The results of moderator testing indicate that objective methodological rigor, as measured by the SMS, significantly influenced arrest outcome for treatment group offenders compared to control group offenders. However, the differences in observed subsequent arrests by different SMS categories do not have a clear pattern. One limitation of this findings is the breakdown of SMS scores. More research is needed to clarify patterns between methodological strength on success outcomes.

Limitations of the Current Study

Additional limitations were present in the current study. The main shortcoming of the current study meta-analyses was a restricted sample, particularly for assessing heterogeneity. Two out of three outcome samples were too small to perform any type of moderator analysis (i.e., victim satisfaction, restitution compliance). It is also possible that the sample that was robust enough for subgroup analysis (i.e., subsequent arrest)

could have been underpowered for detecting moderating effects due to the composition of indicator variable groupings. Additionally, the quality of meta-analyses were dependent on the sampling techniques employed by the author. By design, the sample selection process involved excluding evaluations of programs outside of the United States. While the somewhat limited number of available studies lend support to the need for more evaluative efforts within the United States, it ultimately led to an underpowered quantitative analysis. Conversely, the inclusion of non-published, non-peer-reviewed studies which was meant to avoid publication bias may have confounded results.

The definition of restorative justice used to guide systematic searches may have also been problematic by being overly broad. The intent was to collect all the available evidence on a variety of programs and program features to glean general insights into overall effectiveness of restorative justice in the United States and then to account for any differential outcomes using subgroup analysis. However, there may not have been enough studies in the sample to generate meaningful indicator groupings based on study characteristics. Alternatively, the range in number of studies within distinct indicator variable categories may have been too great to detect moderating effects with certainty (e.g., offender type, offense type).

The obstacles of standardization in success measures for restorative justice abound. Part of the issue lies with over-generalized, broad definitions of program goals. The terms “offender rehabilitation,” “victim recovery,” and “community restoration” are ambiguous and signify different things to different people which diversifies their measurement in empirical research. These goals do not readily lend themselves to a straightforward quantitative analysis, especially for non-recidivistic outcomes.

Qualitative or mixed methods research may better assess the true impacts of restorative justice for all involved stakeholders or provide a framework from which to generate quantitative measurement of qualitative constructs. Mixed methods data collection requires great economic and time resources than may be currently available for research. However, these practical limitations need to be addressed in order to validate the effectiveness of restorative justice programs in terms of its ascribed goals.

Implications for Future Research and Practice

Inherent challenges associated with evaluating the success of restorative justice programs in the United States are difficult to overcome. Nevertheless, identifying avenues from measuring non-traditional outcomes is a needed step in future research which seeks to clarify the effectiveness of these programs. Compared to the current study, a larger meta-analysis sample is needed to better assess the program characteristics and study characteristics that most impact success. This is especially true for evaluating non-recidivism outcomes. Additionally, other potential moderators concerning program fidelity (e.g., use of trained facilitator, program setting, specific offense) may be meaningful, but were not assessed in the current study. In order to provide a global context for the success of restorative justice in the United States future research should include international samples, but account for country of origin in moderator analysis.

Though evaluations have used an array of outcome measures as proxies for programmatic success they have been limited in their inclusion of support and community stakeholders and have not assessed community-level impacts. Identifying ways to operationalize the unique goals of restorative justice is paramount to determining the value of restorative justice programs to benefit those touched by crime. Restorative

justice traditionalists maintain that voluntary participation by all stakeholders is integral to the restorative process. However, this stipulation is not always met by the implementation practices within the United States. Innovative measurement of the construct of voluntary participation of offenders is needed to inform the debate about the negative ramifications of unwilling participants on the victims of crime.

Limitations notwithstanding, the findings of the current study support the use of restorative justice to decrease recidivism and increase victim satisfaction and restitution compliance. These findings encourage the continued use of these programs as an alternative to traditional criminal justice. Implications to practice based on program type, offense type, offender age, and voluntary participation are complex. Due to sample size restrictions only subsequent arrest outcomes could be examined more closely for influences of these indicator categories. Therefore, the potential moderating effects of these groupings on victim satisfaction and restitution compliance remain unknown. Findings among subsequent arrest subgroup analysis support the continued use of all restorative justice programs, with an emphasis on the success of conferencing, hybrid, and circle modalities. Heterogeneity testing also indicates that mandated offender participation, diverse offender types, and use of restorative justice with adults are not associated with poorer outcomes. However, these findings are constrained by skewed distributions of category levels (i.e., majority juvenile participants, mixed crime type, and somewhat voluntary participation). The author suggests that restorative justice continue to be implemented in cases of serious crime and adult offending guided by safe practice protocols such as secure program setting, respectful language, use of trained facilitators, and program preparation of stakeholders. Skepticism of the effectiveness and

appropriateness of restorative justice is warranted, but the true impacts of these programs cannot be assessed without their implementation and continued evaluation.

Conclusion

The need to repair harms caused by crime in the U.S. has historically been overshadowed by the retributive nature of western justice practices. The adoption of restorative justice as an alternative to formal criminal punishment has spurred decades' worth of evaluative research including several meta-analyses, yet this literature has failed to provide clear evidence of program effectiveness. As the use of restorative justice programs increases within America greater empirical attention is needed to explore their true impacts on stakeholders. This investigation aimed to gain a better understanding of the state of evaluative literature on restorative justice programs in the United States. Specifically, the author reviewed the dependent variables operationalized as success, determined study quality of methods, and synthesized research findings in a representative sample of evaluations identified by systematic search protocol. Most salient among research findings was the lack of available research examining outcomes of restorative justice programs in the United States. Although the sample size of subsequent arrest meta-analysis was robust and comparable to previous literature, the limited number of studies reporting on victim satisfaction and restitution compliance hindered the authors ability to quantify success in non-recidivistic ways. Future application of these practices with diverse populations as well as sound evaluative research designs are imperative to determining the full spectrum of restorative justice impacts on stakeholders.

References

*Studies included for coding

**Studies included for both coding and meta-analysis

*Acosta, J. D., Chinman, M., Ebener, P., Phillips, A., Xenakis, L., & Malone, P. S. (2016). A cluster-randomized trial of restorative practices: An illustration to spur high-quality research and evaluation. *Journal of Educational and Psychological Consultation*, *26*(4), 413-430.

Alexander, M. (2012). *The new Jim Crow: Mass incarceration in the age of colorblindness*. The New Press.

Alper, M., Durose, M. R., & Markman, J. (2018). *2018 Update on Prisoner Recidivism: A 9-Year Follow-up Period (2005-2014)*. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

*Armour, M. (2012). Ed. White Middle School Restorative Discipline Evaluation: Implementation, and Impact.

Anders, M. E., & Evans, D. P. (2010). Comparison of PubMed and Google Scholar literature searches. *Respiratory care*, *55*(5), 578-583.

Angel, C. M. (2005). *Crime Victims Meet Their Offenders: Testing the Impact of Restorative Justice Conferences on Victim's Post-traumatic Stress Symptoms* (Doctoral dissertation, University of Pennsylvania).

Anno, B. J. (1990). The cost of correctional health care: Results of a national survey. *Journal of Prison and Jail Health*, *9*(2), 105-133.

Aromataris, E., & Pearson, A. (2014). The systematic review: an overview. *AJN The American Journal of Nursing*, *114*(3), 53-58.

*Baffour, T. D. (2006). Ethnic and gender differences in offending patterns: Examining family group conferencing interventions among at-risk adolescents. *Child and Adolescent Social Work Journal*, *23*(5-6), 557-578.

*Baglivio, M., & Jackowski, K. (2015). Evaluating the effectiveness of a victim impact intervention through the examination of changes in dynamic risk scores. *Criminal Justice Policy Review*, *26*(1), 7-28.

*Baker, M. L. (2008). DPS restorative justice project: Executive summary, 2007-2008.

Bales, W. D., Van Slyke, S., & Blomberg, T. G. (2006). Substance abuse treatment in prison and community reentry: Breaking the cycle of drugs, crime, incarceration, and recidivism. *Geo. J. on Poverty L. & Pol'y*, *13*, 383

Barnett, R. (1977). Restitution: A new paradigm of criminal justice. Ethics, 87, 279-301

Bartels, E. M. (2013). How to perform a systematic search. *Best Practice & Research Clinical Rheumatology*, *27*(2), 295-306.

Bartolucci, A. A., & Hillegass, W. B. (2010). Overview, strengths, and limitations of systematic reviews and meta-analyses. In *Evidence-based practice: Toward optimizing clinical outcomes* (pp. 17-33). Springer, Berlin, Heidelberg.

Bazemore, G. (1996). Three paradigms for juvenile justice. *Restorative justice: International perspectives*, *37*, 67.

Bazemore, G. (1998). Restorative justice and earned redemption communities, victims, and offender reintegration. American Behavioral Scientist, 41(6), 768-813.

Bazemore, G., & Day, S. (1996). Restoring the balance: Juvenile and community justice. *Juv. Just.*, *3*, 3.

Bazemore, G., & Maloney, D. (1994). Rehabilitating community service toward restorative service sanctions in a balanced justice system. *Fed. Probation*, *58*, 24.

Bazemore, S. G., Pranis, K., & Umbreit, M. S. (1997). Balanced and restorative justice for juveniles: A framework for juvenile justice in the 21st century. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Bazemore, G., & Schiff, M. (2013). *Juvenile justice reform and restorative justice*. Routledge.

Bazemore, G., & Umbreit, M. S. (1994). *Balanced and restorative justice*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Bazemore, G., & Umbreit, M. (1995). Rethinking the sanctioning function in juvenile court: Retributive or restorative responses to youth crime. *Crime & Delinquency*, *41*(3), 296-316.

Bazemore, G., & Umbreit, M. (2001). A Comparison of Four Restorative Conferencing Models. *Juvenile Justice Bulletin*.

Bazemore, G., & Schiff, M. (2015). *Restorative community justice: Repairing harm and transforming communities*. Routledge.

Beale, S. S. (2003). Still tough on crime-prospects for restorative justice in the United States. *Utah L. Rev.*, 413.

Beck, R. J. (1997). Communications in a teen court: Implications for probation. *Fed. Probation*, *61*, 40.

Becker, H. S. (1960). Notes on the concept of commitment. *American journal of Sociology*, *66*(1), 32-40.

*Beck-Zierdt, N. (1980). Tri-county juvenile restitution program. St. Paul, MN: Research and Evaluation Unit, Crime Control Planning Board.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088-1101.

**Behtz, S. A. (2004). Justice for All?: Victim Satisfaction with Restorative Justice Conferences.

Belgrave, J. (1995). Restorative Justice: a Discussion Paper. New Zealand Ministry of Justice. Wellington, New Zealand.

Benesh, S. C. (2006). Understanding public confidence in American courts. Journal of Politics, 68(3), 697-707.

Bennett, M. W. (2016). The implicit racial bias in sentencing: The next frontier. *Yale LJF*, *126*, 391.

**Berger, D. E., Lipsey, M. W., Dennison, L. B., & Lange, J. M. (1977). The Effectiveness of the Sheriff's Department's Juvenile Diversion Projects in Southeast Los Angeles County (CA).

**Bergseth, K. J., & Bouffard, J. A. (2007). The long-term impact of restorative justice programming for juvenile offenders. *Journal of Criminal Justice*, *35*(4), 433-451.

**Bergseth, K. J., & Bouffard, J. A. (2013). Examining the effectiveness of a restorative justice program for various types of juvenile offenders. *International journal of offender therapy and comparative criminology*, *57*(9), 1054-1075.

Bernburg, J. G. (2009). Labeling theory. In *Handbook on crime and deviance* (pp. 187-207). Springer, New York, NY.

Bero, L. A., Grilli, R., Grimshaw, J. M., Harvey, E., Oxman, A. D., & Thomson, M. A. (1998). Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *Bmj*, *317*(7156), 465-468.

Bibas, S., & Bierschbach, R. A. (2004). Integrating remorse and apology into criminal procedure. *Yale LJ*, *114*, 85.

Bigby, M. (2014). Understanding and evaluating systematic reviews and meta-analyses. *Indian journal of dermatology*, *59*(2), 134.

Bilz, K. (2007). The puzzle of delegated revenge. *BUL Rev.*, *87*, 1059.

Bloch, K. E. (2010). Reconceptualizing restorative justice. *Hastings Race & Poverty LJ*, *7*, 201.

Blumstein, A., & Larson, R. C. (1971). Problems in modeling and measuring recidivism. *Journal of Research in Crime and Delinquency*, *8*(2), 124-132.

Bonczar, T., Hughes, T., Wilson, D., Ditton, P. (2018). National Corrections Reporting Program. *Bureau of Justice Statistics.*

Bonta, J., Jesseman, R., Rugge, T., & Cormier, R. (2006). Restorative justice and recidivism: Promises made, promises kept. *Handbook of restorative justice: A global perspective*, *108*, 120.

Bonta, J., Rugge, T., Scott, T. L., Bourgon, G., & Yessine, A. K. (2008). Exploring the black box of community supervision. *Journal of offender rehabilitation*, *47*(3), 248-270.

Bonta, J., Rooney, J., & Wallace-Capretta, S. M. (1998). *Restorative justice: An evaluation of the restorative resolutions project*. Ottawa: Solicitor General Canada.

Bonta, J., Wallace-Capretta, S., Rooney, J., & McAnoy, K. (2002). An outcome evaluation of a restorative justice alternative to incarceration. *Contemporary Justice Review*, *5*(4), 319-338.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2018). Comprehensive Meta-Analysis Version 3. Biostat, Englewood, NJ 2013

*Bouffard, J., Cooper, M., & Bergseth, K. (2017). The effectiveness of various restorative justice interventions on recidivism outcomes among juvenile offenders. *Youth violence and juvenile justice*, *15*(4), 465-480.

Bradshaw, W., Roseborough, D., & Umbreit, M. S. (2006). The effect of victim offender mediation on juvenile offender recidivism: A meta-analysis. *Conflict Resolution Quarterly*, *24*(1), 87-98.

*Bradshaw, W., & Umbreit, M. S. (1998). Crime victims meet juvenile offenders: Contributing factors to victim satisfaction with mediated dialogue. *Juvenile and Family Court Journal*, *49*(3), 17-25.

Braithwaite, J. (1989). *Crime, shame and reintegration*. Cambridge University Press.

Braithwaite, J. (1999). Restorative justice: Assessing optimistic and pessimistic accounts. *Crime and justice*, *25*, 1-127.

Braithwaite, J. (2004). Restorative justice: Theories and worries. *Resource Material Series*, *63*, 47-56.

Braithwaite, J., & Mugford, S. (1994). Conditions of successful reintegration ceremonies: Dealing with juvenile offenders. *The British journal of criminology*, *34*(2), 139-171.

Braithwaite, J., & Roche, D. (2001). Responsibility and restorative justice. *Restorative community justice: repairing harm and transforming communities. Cincinnati: Anderson*, 63-84.

Braman, D. (2007). *Doing time on the outside: Incarceration and family life in urban America*. University of Michigan Press.

**Brooks, A. (2013). Moving forward: Two approaches to repairing the harm through restorative justice. American University.

Brown, D. (2010). The limited benefit of prison in controlling crime. *Current Issues in Criminal Justice*, *22*(1), 137-148.

Brunk, C. (2001). Restorative justice and the philosophical theories of criminal punishment. *The spiritual roots of restorative justice*, 31-56.

Buchanan, K. S. (2007). Impunity: Sexual abuse in women's prisons. *Harv. CR-CLL Rev.*, *42*, 45.

Burke, A. (2007). Prosecutorial passion, cognitive bias, and plea bargaining. *Marquette Law Review*, *93*, 183–211.

Bushie, B. (1997). A personal journey. *The four circles of hollow water*.

*Butts, J., Buck, J., & Coggeshall, M. (2002). The Impact of Teen Court on Young Offenders. Research Report.

*Butts, J. & Snyder, H. (1992). Restitution and Juvenile Recidivism (NCJRS No. 137774). U. S. Department of Justice, Washington, DC.

Byrd, J. (2008). J. Mitchell Miller, Christopher L. Gibson and. *Restorative Justice: From Theory to Practice*, *11*, 261-278.

*Cannon, A., & Stanford, R. M. (1981). Evaluation of the Juvenile Alternative Services Project (NCJRS No. 080633). Tallahassee, FL: Office of Children, Youth, and Families.

**Carr, C., & Nelson, P. (2000). Centinela Valley's victim offender restitution services: A report for the administrative office of the courts, judicial council of California. Los Angeles, California. Chapter 3 from Evje and Cushman Book

Casper, J. D., Tyler, T., & Fisher, B. (1988). Procedural justice in felony cases. *Law & Soc'y Rev.*, *22*, 483.

Cayley, D. (1998). *The expanding prison: The crisis in crime and punishment and the search for alternatives*. House of Anansi.

Cerrato, S. (2014). Achieving reform in unstable correctional institutions: a theoretical perspective–revisited. *Contemporary Justice Review*, 17(2), 273-296.

Chamberlain, A. W., & Hipp, J. R. (2015). It's all relative: Concentrated disadvantage within and across neighborhoods and communities, and the consequences for neighborhood crime. *Journal of Criminal Justice*, *43*(6), 431-443.

Chappell, A. T. (2018). Predicting the Behavior of Law in the Juvenile Court: A Focus on Noncompliance Cases. *Crime & Delinquency*, 0011128718787156.

Chiricos, T., Barrick, K., Bales, W., & Bontrager, S. (2007). The labeling of convicted felons and its consequences for recidivism. *Criminology*, *45*(3), 547-581.

Choi, J. J., Bazemore, G., & Gilbert, M. J. (2012). Review of research on victims' experiences in restorative justice: Implications for youth justice. *Children and Youth Services Review*, *34*(1), 35-42.

Choi, J. J., & Severson, M. (2009). "What! What kind of apology is this?": The nature of apology in victim offender mediation. *Children and youth services review*, *31*(7), 813-820.

*Clarke, S. H., Valente, E., & Mace, R. R. (1992). Mediation of interpersonal disputes: An evaluation of North Carolina's programs. Chapel Hill, NC: Institute of Government, University of North Carolina at Chapel Hill.

Clear, T. R. (2009). *Imprisoning communities: How mass incarceration makes disadvantaged neighborhoods worse*. Oxford University Press.

*Coates, R. B., & Gehm, J. (1985). Victim meets offender: An evaluation of victim-offender reconciliation programs. PACT Institute of Justice.

*Coates, R., Umbreit, M., & Vos, B. (2003). Restorative justice circles: An exploratory study. Contemporary Justice Review, 6(3), 265-278.

*Coates, R. B., Vos, B., & Umbreit, M. (2000). Restorative Justice Circles in South Saint Paul, Minnesota. University of Minnesota. Center for Restorative Justice & Peacemaking.

Cochran, J. C. (2014). Breaches in the wall: Imprisonment, social support, and recidivism. *Journal of research in crime and delinquency*, *51*(2), 200-229.

*Coldren Jr, J. R., Haring, C., Luecke, A., Sintic, C., & Balgoyen, S. (2011). School-based Restorative Justice Data Template.

Consedine, J. (1995). *Restorative justice: Healing the effects of crime* (p. 176). Lyttelton, New Zealand: Ploughshares Publications.

Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of internal medicine, 126*(5), 376-380.

Cook, D. A., & West, C. P. (2012). Conducting systematic reviews in medical education: a stepwise approach. *Medical education, 46*(10), 943-952.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological bulletin*, *87*(3), 442.

**Cosden, M., Casas, M., & Wolfe, M. (1999). Evaluation of Santa Barbara's restorative justice project. Santa Barbara: Counseling, Clinical, School Psychology Program, University of California, Santa Barbara, from Cushman, R. C. (2000). A Summary of the Evaluations of Six California Victim Offender Reconciliation Programs.

Cox, G. H., & Rhodes, S. L. (1990). Managing Overcrowding: Corrections Administrators and The Prison Crisis. *Criminal Justice Policy Review, 4*(2), 115-143.

**Crofoot, J. A. (1988). A Juvenile Diversion Program's Effectiveness with Varying Levels of Offender Severity.

*Crotty, J., Meier, R. D., Behavioral Systems Associates, Inc, United States of America, Thamens Valley Council for Community Action, Inc, & United States of America. (1980). Evaluation of juvenile restitution program Project Detour: Final report. East Lyme, CT: Behavioral Systems Associates.

Cullen, F. T. (1995). Assessing the penal harm movement. *Journal of research in crime and delinquency*, 32, 338-358.

Cullen, F. T. (2005). The twelve people who saved rehabilitation: how the science of criminology made a difference: the American Society of Criminology 2004 presidential address. Criminology, 43(1), 1-42.

Cullen, F. T. (2013). Rehabilitation: Beyond nothing works. *Crime and Justice, 42*(1), 299-376.

Cullen, F. T., & Gendreau, P. (2001). From nothing works to what works: Changing professional ideology in the 21st century. *The Prison Journal, 81*(3), 313-338.

Cullen, F. T., & Gilbert, K. E. (2012). *Reaffirming rehabilitation*. Routledge.

Cullen, F. T., Jonson, C. L., & Nagin, D. S. (2011). Prisons do not reduce recidivism: The high cost of ignoring science. *The Prison Journal, 91*(3), 48-65.

Cullen, F. T., Skovron, S. E., Scott, J. E., & Burton Jr., V. S. (1990). Public support for correctional treatment: The tenacity of rehabilitative ideology. *Criminal Justice and Behavior, 17*(1), 6-18.

Currie, E. (2013). *Crime and punishment in America*. Macmillan.

Daly, K. (2016). What is restorative justice? Fresh answers to a vexed question. Victims & Offenders, 11(1), 9-29.

Daly, K. (2017). Restorative justice: The real story. In *Restorative Justice* (pp. 85-109). Routledge.

**Davis, R. C. (2009). The Brooklyn mediation field test. *Journal of experimental criminology, 5*(1), 25-39

**De Beus, K., & Rodriguez, N. (2007). Restorative justice practice: An examination of program completion and recidivism. *Journal of Criminal Justice, 35*(3), 337-347.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, *7*(3), 177-188.

Devers, L. (2011). Plea and charge bargaining. *Research summary*, *1*.

*DeWitt, D. M., & DeWitt, L. J. (2012). A case of high school hazing: Applying restorative justice to promote organizational learning. *NASSP Bulletin, 96*(3), 228-242.

**Dick, E. Victim Offender Reconciliation Program of Mendocino County. VORP of Mendocino County (Mendocino, Calif., November 1999), from Cushman, R. C. (2000). A Summary of the Evaluations of Six California Victim Offender Reconciliation Programs.

Dodson, K. D., Cabage, L. N., & Klenowski, P. M. (2011). An evidence-based assessment of faith-based programs: Do faith-based programs "work" to reduce recidivism?. *Journal of Offender Rehabilitation*, *50*(6), 367-383.

Dooley, M. J. (1995). Reparative probation program. *Monograph. Vermont Department of Corrections*.

Duval, S. J. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.) *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 127–144). Chichester, England: Wiley.

Duval, S. J., & Tweedie, R. L. (2000a). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.

Duval, S. J., & Tweedie, R. L. (2000b). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98.

Duwe, G. (2009). Residency restrictions and sex offender recidivism: Implications for public safety. *Geography & Public Safety*, *2*(1), 6-8.

Duwe, G. (2013). Can Circles of Support and Accountability (COSA) work in the United States? Preliminary results from a randomized experiment in Minnesota. *Sexual Abuse*, *25*(2), 143-165.

**Duwe, G. (2018). Can circles of support and accountability (CoSA) significantly reduce sexual recidivism? Results from a randomized controlled trial in Minnesota. *Journal of Experimental Criminology*, *14*(4), 463-484.

Egger, M., Dickersin, K., & Smith, G. D. (2001). Problems and limitations in conducting systematic reviews. *Systematic reviews in health care: Meta-analysis in context*, 43-68.

Egger, M., Smith, G. D., & Phillips, A. N. (1997). Meta-analysis: principles and procedures. *Bmj*, *315*(7121), 1533-1537.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629-634.

Eglash, A. (1958). Creative restitution: its roots in psychiatry, religion and law. Brit. J. Delinq., 10, 114.

Engen, R. L., & Steen, S. (2000). The power to punish: Discretion and sentencing reform in the war on drugs. *American Journal of Sociology*, *105*(5), 1357-1395.

Everett, R. S., & Wojtkiewicz, R. A. (2002). Difference, disparity, and race/ethnic bias in federal sentencing. *Journal of Quantitative Criminology*, *18*(2), 189-211.

Everett, R. S., & Wojtkiewicz, R. A. (2002). Difference, disparity, and race/ethnic bias in federal sentencing. *Journal of Quantitative Criminology*, *18*(2), 189-211.

*Ezell, R. (1986). Juvenile Arbitration as a Diversionary Alternative (Net-widening, social control).

Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2002). The Maryland scientific methods scale. *Evidence-based crime prevention*, 13-21.

Farrington, D. P., & Murray, J. (2013). *Labeling Theory: Empirical Tests.* New Brunswick.

*Featherston, T. R. (2014). An experimental study on the effectiveness of a restorative justice intervention on the social aggression, social problem solving skills, and prosocial behaviors of African American adolescent girls (Doctoral dissertation, Capella University).

Federal Bureau of Prisons. (2019, April 13). *Inmate Statistics, Offenses.* Retrieved from http://www.bop.org

Feeley, M. (2017). Two models of the criminal justice system: An organizational perspective. In *Crime, Law and Society* (pp. 119-137). Routledge.

Feeley, M., & Simon, J. (1992) The new penology: notes on the emerging strategy of corrections and its implications. *Criminology* 30, 449–474

*Fercello, C., & Umbreit, M. (1999). Client satisfaction with victim-offender conferences in Dakota County, Minnesota. Center for Restorative Justice and Peacemaking, University of Minnesota, St. Paul.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed-and random-effects methods. *Psychological methods*, *6*(2), 161.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 665-694.

*Fishbein, P., Davis, J. M., & Hamparin, D. (1984). Restitution programming for juvenile offenders. Columbus, OH: Ohio Serious Juvenile Offender Program, Department of Corrections.

*Flaten, C. (1996). Victim-offender mediation: Application with serious offenses committed by juveniles. *Restorative justice: International perspectives*, 387-402.

Fogel, D., Galaway, B., & Hudson, J. (1972). Restitution in criminal justice: A Minnesota experiment. *Criminal Law Bulletin*, *8*(8), 681-91.

*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.

**Fors, S. W., & Rojek, D. G. (1999). The effect of victim impact panels on DUI/DWI rearrest rates: a twelve-month follow-up. *Journal of Studies on Alcohol*, *60*(4), 514-520.

*Fox, K. J. (2013). Circles of support & accountability: Final report prepared for the State of Vermont Department of Corrections. Retrieved from: http://www.doc.state.vt.us/about/reports/circles-of-supportaccountability- final-report/view.

Fox, K.A., & Shjarback, J. (2016). What works to reduce victimization? A systematic evaluation. *Violence & Victims, 31,* 285-319.

Freeman, R. (2003). *Can we close the revolving door?: Recidivism vs. employment of ex-offenders in the US*. Urban Institute.

*Freeman, M. (2018). Impact of Restorative Justice on 11th and 12th Grade African American Male Student Success in a Midwestern City Large Urban Charter School District (Doctoral dissertation, University of St. Francis).

Fronius, T., Persson, H., Guckenburg, S., Hurley, N., & Petrosino, A. (2016). Restorative justice in US schools: A research review. *San Francisco, CA: WestEd Justice and Prevention Training Center*.

**Fulkerson, A. (2001). The use of victim impact panels in domestic violence justice approach. *Contemporary Justice Review*, 4, 355-368.

*Galaway, B. (1988). Crime victim and offender mediation as a social work strategy. *Social Service Review, 62*(4), 668-683.

Galaway, B., & Hudson, J. (Eds.). (1990). *Criminal justice, restitution, and reconciliation*. Monsey, NY: Criminal Justice Press.

Gau, J. M. (2015). Procedural justice, police legitimacy, and legal cynicism: A test for mediation effects. *Police Practice and Research*, *16*(5), 402-415.

Gaes, G. G., Wallace, S., Gilman, E., Klein-Saffran, J., & Suppa, S. (2002). The influence of prison gang affiliation on violence and other prison misconduct. *The Prison Journal*, *82*(3), 359-385.

**Gase, L. N., Kuo, T., Lai, E. S., Stoll, M. A., & Ponce, N. A. (2016). The impact of two Los Angeles County Teen Courts on youth recidivism: comparing two informal probation programs. *Journal of experimental criminology, 12*(1), 105-126.

**Griffith, W. R., Institute of Policy Analysis, & United States of America. (1983). The effect of Washington DC's restitution program on the recidivism rates of the disadvantaged, serious offender. NCJ, (098581).

Gilbert, M. J., Schiff, M., & Cunliffe, R. H. (2013). Teaching restorative justice: developing a restorative andragogy for face-to-face, online and hybrid course modalities. *Contemporary justice review*, *16*(1), 43-69.

Goldkamp, J. S. (2003). The impact of drug courts. *Criminology & Public Policy*, *2*(2), 197-206.

Greenwood, P. W., & Turner, S. (2011). Juvenile crime and juvenile justice. *Crime and public policy*, 88-129.

Gross, J. (2010). The effects of net-widening on minority and indigent drug offenders: A critique of drug courts. *U. Md. LJ Race, Religion, Gender & Class*, *10*, 161.

*Guedalia, L. J. (1979). Predicting recidivism of juvenile delinquents on restitutionary probation from selected background, subject and program variables (Doctoral dissertation, The American University).

*Haarman, G.B., & Covington, C. (1981). Juvenile restitution project: An evaluation. Jefferson County, KY: Department of Human Services, Office of Research and Planning.

Harris, M. K. (2004). An expansive, transformative view of restorative justice. *Contemporary Justice Review*, *7*(1), 117-141.

Harris, N., & Burton, J. B. (1998). Testing the reliability of observational measures of reintegrative shaming at community accountability conferences and at court. *Australian & New Zealand Journal of Criminology*, *31*(3), 230-241.

Harris, P. W., Lockwood, B., Mengers, L., & Stoodley, B. (2011). Measuring recidivism in juvenile corrections. *Journal of Juvenile Justice*, *1*(1).

Harris, N., Walgrave, L., & Braithwaite, J. (2004). Emotional dynamics in restorative conferences. *Theoretical criminology*, *8*(2), 191-210.

Harrison, L. D. (2001). The revolving prison door for drug-involved offenders: Challenges and opportunities. *Crime & delinquency*, *47*(3), 462-485.

*Harrison, P., Maupin, J. R., & Mays, G. L. (2001).Teen Court: An examination of the processes and outcomes. *Crime & Delinquency*, 47, 243-264.

Hay, C. (2001). An exploratory test of Braithwaite's reintegrative shaming theory. *Journal of Research in Crime and Delinquency*, *38*(2), 132-153.

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods*, *3*(4), 486.

*Heinz, J., Galaway, B. and Hudson, J. (1976). Restitution or parole: A follow-up study of adult offenders. *Social Service Review*, 50, 148–156.

*Helfgott, J. B., Lovell, M. L., Lawrence, C. F., & Parsonage, W. H. (2000). Results from the pilot study of the citizens, victims, and offenders restoring justice program at the Washington state reformatory. *Journal of Contemporary Criminal Justice*, *16*(1), 5-31.

Henrichson, C., & Delaney, R. (2012). The price of prisons: What incarceration costs taxpayers. *Fed. Sent'g Rep.*, *25*, 68.

**Henggeler, S. W., Melton, G. B., & Smith, L. A. (1992). Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology, 60*(6), 953.

Higgins, J. P., & Green, S. (Eds.). (2008). Cochrane handbook for systematic reviews of interventions.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, *21*(11), 1539-1558.

Hillsman, S., & Greene, J. A. (1992). The use of fines as an intermediate sanction. *Smart sentencing: The emergence of intermediate sanctions*, 123-141.

**Hitao, G. (1999). Chapter 8: The VORP in Sonoma County. In A. Evje & R. C. Cushman (Eds.), A summary of the evaluations of six California victim offender reconciliation programs (pp. 91-100). San Francisco: The Judicial Council of California, Administrative Office of the Courts, Center for Families, the Children, & the Courts.

Hotaling, G. T., & Buzawa, E. S. (2003). *Victim satisfaction with criminal justice case processing in a model court setting*. Department of Criminal Justice, University of Massachusetts, Lowell.

Hoffman, M. B. (2000). The drug court scandal. North Carolina Law Review, 78, 1439-1534

Hoffman, M. B. (2017). The Denver drug court and its unintended consequences. In *Drug Courts* (pp. 67-87). Routledge.

Hudson, J., & Galaway, B. (1975). *Considering the victim: Readings in restitution and victim compensation*. Charles C. Thomas Publisher.

Hughes, P., & Mossman, M. J. (2001). *Re-Thinking access to criminal justice In Canada: a critical review of needs, responses and restorative justice initiatives*. Research and Statistics Division, Department of Justice Canada.

*Hughes, S. P., & Schneider, A. L. (1989). Victim-offender mediation: A survey of program characteristics and perceptions of effectiveness. *Crime & Delinquency*, *35*(2), 217-233.

*Hughes, S. & Schneider, A. (1990). Victim-Offender Mediation in the Juvenile Justice System. Washington, D.C.: Office of Juvenile Justice and Delinquency Prevention.

Iguchi, M. Y., Bell, J., Ramchand, R. N., & Fain, T. (2005). How criminal system racial disparities may translate into health disparities. *Journal of health care for the poor and underserved*, *16*(4), 48-56.

Immarigeon, R. (1999). Restorative justice, juvenile offenders and crime victims: A review of the literature. *Bazemore and Walgrave, eds*, *305*.

*Jackson, A. L. (2009). The impact of restorative justice on the development of guilt, shame, and empathy among participants in a victim impact training program. *Victims and Offenders*, *4*(1), 1-24.

Jenson, J. (2010). *Defining and measuring social cohesion* (No. 1). Commonwealth Secretariat.

**Jeong, S., McGarrell, E. F., & Hipple, N. K. (2012). Long-term impact of family group conferences on re-offending: The Indianapolis restorative justice experiment. *Journal of Experimental Criminology*, *8*(4), 369-385.

Johnstone, G. (2013). Restorative justice: Ideas, values, debates. Routledge.

Johnstone, G., & Van Ness, D. (2004). The idea of restorative justice. Inaugural Lecture at The Middleton Hall, University of Hull.

Karp, D. R. (2001). Harm and repair: Observing restorative justice in Vermont. *Justice Quarterly*, *18*(4), 727-757.

*Karp, D. R., & Drakulich, K. M. (2004). Minor crime in a quaint setting: Practices, outcomes, and limits of Vermont reparative probation boards. *Criminology & Public Policy*, *3*(4), 655-686.

Kaukinen, C. (2004). The help-seeking strategies of female violent-crime victims: The direct and conditional effects of race and the victim-offender relationship. *Journal of interpersonal violence*, *19*(9), 967-990.

Khadjavi, M. (2018). Deterrence works for criminals. *European Journal of Law and Economics*, *46*(1), 165-178.

Kim, B., Cano, M. V., Kim, K., & Spohn, C. (2016). The impact of United States v. Booker and Gall/Kimbrough v. United States on sentence severity: Assessing social context and judicial discretion. *Crime & Delinquency*, *62*(8), 1072-1094.

King, N., Soule, D., Steen, S., & Weidner, R. (2005). When process affects punishment: Differences in sentences after guilty plea, bench trial, and jury trial in five guideline states. *Columbia Law Review, 105*, 960–1009.

Kirchengast, T. (2016). *The victim in criminal law and justice*. Springer.

Kirk, D. S., & Papachristos, A. V. (2011). Cultural mechanisms and the persistence of neighborhood violence. *American journal of sociology*, *116*(4), 1190-1233.

Kirk, D. S., & Papachristos, A. V. (2017). Concentrated disadvantage, the persistence of legal cynicism, and crime: Revisiting the conception of "culture" in criminology. In *Challenging Criminological Theory* (pp. 259-274). Routledge.

Klein, A. R. (1997). *Alternative sentencing, intermediate sanctions, and probation*. Cincinnati, OH: Anderson.

*Koch, R. (1986). Community service and outright release as alternatives to juvenile court: An experimental evaluation (Doctoral dissertation).

*Koss, M. P. (2014). The RESTORE program of restorative justice for sex crimes: Vision, process, and outcomes. *Journal of Interpersonal Violence, 29*(9), 1623-1660.

Koss, M. P., Bachar, K. J., & Hopkins, C. Q. (2003). Restorative justice for sexual violence. *Annals of the New York Academy of Sciences*, *989*(1), 384-396.

Kramer, J., & Steffensmeir, D. (1993). Race and imprisonment decisions. *The Sociological Quarterly*, *34*(2), 357-376.

Kunst, M., Popelier, L., & Varekamp, E. (2015). Victim satisfaction with the criminal justice system and emotional recovery: A systematic and critical review of the literature. *Trauma, Violence, & Abuse*, *16*(3), 336-358.

Kubrin, C. E., & Stewart, E. A. (2006). Predicting who reoffends: The neglected role of neighborhood context in recidivism studies. *Criminology*, *44*(1), 165-197.

Kurlychek, M. C., & Johnson, B. D. (2004). The juvenile penalty: A comparison of juvenile and young adult sentencing outcomes in criminal court. *Criminology*, *42*(2), 485-515.

Kurki, L. (2000). Restorative and community justice in the United States. *Crime and justice*, *27*, 235-303.

Kyckelhahn, T. (2011). Justice expenditures and employment, FY 1982–2007–statistical tables. *Bureau of Justice Statistics. Washington, DC: US Department of Justice*.

LaFree, G. (2018). *Losing legitimacy: Street crime and the decline of social institutions in America*. Routledge.

**Lane, J., Turner, S., Fain, T., & Sehgal, A. (2005). Evaluating an experimental intensive juvenile probation program: Supervision and official outcomes. *Crime & Delinquency*, 51, 26-52.

*Lane, J., Turner, S., Fain, T., & Sehgal, A. (2007). The effects of an experimental intensive juvenile probation program on self-reported delinquency and drug use. *Journal of Experimental Criminology*, *3*(3), 201-219.

Latimer, J., Dowden, C., & Muise, D. (2005). The effectiveness of restorative justice practices: A meta-analysis. *The prison journal*, *85*(2), 127-144.

**Lee, S. (1990). Victim Offender Mediation Program Evaluation. Prepared by Community Crime Prevention Associates for the Santa Clara County Probation Department and Santa Clara County Office of Human Relations (San Jose, Calif., December).

Lens, K. M., Pemberton, A., Brans, K., Braeken, J., Bogaerts, S., & Lahlah, E. (2015). Delivering a Victim Impact Statement: Emotionally effective or counter-productive?. *European Journal of Criminology*, *12*(1), 17-34.

*Leslie, B. (2002). The neighborhood response team: community based conflict resolution to promote civic engagement through restorative community justice (Doctoral dissertation, Southern New Hampshire University).

*Levi, K. (1982). Relative redemption: Labeling in juvenile restitution. *Juvenile and Family Court Journal*, 33, 3–13.

*Lewis, S. (2009). Improving School Climate: Findings from Schools Implementing Restorative Practices. International Institute for Restorative Practices Graduate School.

Liles, W. W. (2006). Challenges to felony disenfranchisement laws: Past, present, and future. *Ala. L. Rev.*, *58*, 615.

**Lincoln, S. B., Teilmann, K. S., Klein, M. W., & Labin, S. (1977, February). Recidivism rates of diverted juvenile offenders. In National Conference on Criminal Justice Evaluation, Washington, DC.

Link, B. G., Cullen, F. T., Struening, E., Shrout, P. E., & Dohrenwend, B. P. (1989). A modified labeling theory approach to mental disorders: An empirical assessment. *American sociological review*, 400-423.

*Lipsey, M. W., Cordray, D. S., & Berger, D. E. (1981). Evaluation of a juvenile diversion program: Using multiple lines of evidence. *Evaluation Review*, *5*(3), 283-306.

Lipsey, M. W., & David, B. (2001). *Practical meta-analysis/Mark W. Lipsey and David B. Wilson* (No. 300.72 L5.).

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.

Listwan, S. J., Jonson, C. L., Cullen, F. T., & Latessa, E. J. (2008). Cracks in the penal harm movement: Evidence from the field. *Criminology & Public Policy*, *7*(3), 423-465.

Lum, C., Koper, C. S., & Telep, C. W. (2011). The evidence-based policing matrix. *Journal of Experimental Criminology*, *7*(1), 3-26.

Luna, E. (2003). Punishment theory, holism, and the procedural conception of restorative justice. *Utah L. Rev.*, 205.

Lynch, J. P., & Sabol, W. J. (2004). Assessing the effects of mass incarceration on informal social control in communities. *Criminology & Public Policy*, *3*(2), 267-294.

Madaleno, M., & Waights, S. (2010). Guide to scoring methods using the Maryland Scientific Methods Scale. *London: ESRC Economic and Social Research Council*.

Mair, G. (1991). What works—Nothing or Everything? Measuring the Effectiveness of Sentences. *Home Office Research Bulletin*, *30*, 3-8.

Makkai, T., & Braithwaite, J. (1994). Reintegrative shaming and compliance with regulatory standards. *Criminology*, *32*(3), 361-385.

Maltz, M. D. (1984). *Recidivism*. Michael Maltz.

Marshall, T. F., Fairhead, S., Kingsley, S., & Murphy, D. (1985). *Alternatives to criminal courts: the potential for non-judicial dispute settlement*. Brookfield, VT: Gower.

Marshall, T. F. (1996). The evolution of restorative justice in Britain. *European Journal on Criminal Policy and Research*, *4*(4), 21-43.

Martinson, R. (1974). What works?-Questions and answers about prison reform. *The public interest*, *35*, 22.

Mauer, M. (2018). Long-Term Sentences: Time to Reconsider the Scale of Punishment. *UMKC L. Rev.*, *87*, 113.

*McCold, P. (1998, November). Police facilitated restorative conferencing: What the data show. In Second Annual International Conference on Restorative Justice for Juveniles. Fort Lauderdale, FL: Florida Atlantic University.

McCold, P. (2000). Toward a mid-range theory of restorative criminal justice: A reply to the Maximalist model. *Contemporary Justice Review*, *3*(4), 357-414.

McCold, P. (2004). Paradigm muddle: The threat to restorative justice posed by its merger with community justice. *Contemporary Justice Review*, *7*(1), 13-35.

McCold, P., & Wachtel, T. (2002). Restorative justice theory validation. *Restorative justice: Theoretical foundations*, 110-142.

McCold, P., & Wachtel, T. (2003). In pursuit of paradigm: A theory of restorative justice. In *Paper presented at the XIII World Congress of Criminology* (Vol. 10, p. 15).

McDonald, D., & Carlson, K. E. (1993). *Sentencing in the Federal Courts: Does Race Matter?: the Transition to Sentencing Guidelines, 1986-90: Summary*. DIANE Publishing.

**McGarrell, E. F., & Hipple, N. K. (2007). Family group conferencing and re-offending among first-time juvenile offenders: the Indianapolis experiment. *Justice Quarterly*, *24*(2), 221-246.

**McGarrell, E. F., Olivares, K., Crawford, K., & Kroovand, N. (2000). The Indianapolis Juvenile Restorative Justice Experiment.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica: Biochemia medica, 22(3), 276-282.

*McMorris, B.J., Beckman, K.J., Shea, G., Baumgartner, J., & Eggert, R.C. (2013). Applying restorative justice practices to Minneapolis Public Schools students recommended for possible expulsion. *Cell*, 763, 232-3442.

Mika, H., Achilles, M., Halbert, E., & Amstutz, L. S. (2004). Listening to victims-A critique of restorative justice policy and practice in the United States. *Fed. Probation*, *68*, 32.

Mika, H., & Zehr, H. (2003). A restorative framework for community justice practice. Criminology, conflict resolution and restorative justice, 135-52.

Miller, J., Gibson, C. L., & Byrd, J. (2008). Getting beyond the liberal feel-good: Toward an accountability-based theoretical research program for restorative justice. In *Restorative Justice: from Theory to Practice* (pp. 261-278). Emerald Group Publishing Limited.

**Mills, L. G., Barocas, B., & Ariel, B. (2013). The next generation of court-mandated domestic violence treatment: A comparison study of batterer intervention and restorative justice programs. *Journal of Experimental Criminology*, *9*(1), 65-90.

*Minor, K. I., Wells, J. B., Soderstrom, I. R., Bingham, R., & Williamson, D. (1999). Sentence completion and recidivism among juveniles referred to teen courts. *Crime & Delinquency*, *45*(4), 467-480.

Mitchell, O., Wilson, D. B., Eggers, A., & MacKenzie, D. L. (2012). Assessing the effectiveness of drug courts on recidivism: A meta-analytic review of traditional and non-traditional drug courts. *Journal of Criminal Justice*, *40*(1), 60-71.

Montori, V. M., Wilczynski, N. L., Morgan, D., & Haynes, R. B. (2005). Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *Bmj*, *330*(7482), 68.

Moore, L. L. D. (2001). Disability and illicit drug use: An application of labeling theory. *Deviant Behavior*, *22*(1), 1-21

Morin, K. M. (2013). “Security here is not safe”: violence, punishment, and space in the contemporary US penitentiary. *Environment and Planning D: Society and Space*, *31*(3), 381-399.

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International journal of qualitative methods*, *1*(2), 13-22.

**Myers, W. C., Burton, P. R., Sanders, P. D., Donat, K. M., Cheney, J., Fitzpatrick, T. M., & Monaco, L. (2000). Project Back-on-Track at 1 year: A delinquency treatment program for early-career juvenile offenders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(9), 1127-1134.

Nagel, I. H. (1989). Structuring sentencing discretion: The new federal sentencing guidelines. *J. Crim. L. & Criminology*, *80*, 883.

Nagin, D. S., Cullen, F. T., & Jonson, C. L. (2009). Imprisonment and reoffending. *Crime and justice*, *38*(1), 115-200.

*Nelson, S. (2000). Evaluation of the Restorative Justice Program. Eugene, OR: Lane County Department of Youth Services.

*Newland, G. (1981). A Comparison of Three Delinquency Prevention Alternatives for Juvenile Status Offenders.

*Niemeyer, M., & Shichor, D. (1996). A preliminary study of a large victim/offender reconciliation program. *Fed. Probation*, *60*, 30.

Nolan Jr, J. L. (2003). Redefining criminal courts: Problem-solving and the meaning of justice. *Am. Crim. L. Rev.*, *40*, 1541.

*Norris, A. N. (2008). An exploratory study on the effect of restorative justice on school success and disciplinary incidents. Michigan State University. School of Criminal Justice.

*Northcutt Bohmert, M., Duwe, G., & Kroovand Hipple, N. (2016). Evaluating restorative justice circles of support and accountability: Can social support overcome structural barriers? *International Journal of Offender Therapy and Comparative Criminology*.

*Norton, M. H., Gold, E., & Peralta, R. (2013). Youth Courts and Their Educational Value: An Examination of Youth Courts in Chester, Pennsylvania. Research for Action.

*Nugent, W. R., & Paddock, J. B. (1995). The effect of victim-offender mediation on severity of reoffense. *Mediation Quarterly*, *12*(4), 353-367.

*Nugent, W. R., & Paddock, J. B. (1996). Evaluating the effects of a victim-offender reconciliation program on reoffense. *Research on Social Work Practice, 6*(2), 155-178.

Nugent, W. R., Williams, M., & Umbreit, M. S. (2003). Participation in Victim-Offender Mediation and the prevalence and severity of subsequent delinquent behavior: A meta-analysis. *Utah L. Rev.*, 137.

Office of Management Budget. (2018, May 23). *Budget of the U.S. Government: President's budget FY2018*. Retrieved from https://www.govinfo.gov/app/collection/budget/2018

Ogletree Jr, C. J. (1987). The Death of Discretion Reflections on the Federal Sentencing Guidelines. *Harv. L. Rev.*, *101*, 1938.

Ogletree Jr, C. J. (1995). An essay on the new public defender for the 21st century. *Law & Contemp. Probs.*, *58*, 81.

Palmer, T. (1975). Martinson revisited. *Journal of research in crime and delinquency*, *12*(2), 133-152.

Parhar, K., & Wormith, J. S. (2013). Risk factors for homelessness among recently released offenders. *Journal of Forensic Social Work*, *3*(1), 16-33.

Parsons, J., & Bergin, T. (2010). The impact of criminal justice involvement on victims' mental health. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, *23*(2), 182-188.

Patrick, S. (1998). Differences in inmate-inmate and inmate-staff altercations: Examples from a medium security prison. *The Social Science Journal*, *35*(2), 253-263.

**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*(1), 129-135.

Pavelka, S. (2016). Restorative justice in the states: An analysis of statutory legislation and policy. *Justice Policy Journal*, *2*(13), 1-23.

Petersilia, J., & Greenwood, P. W. (1978). Mandatory prison sentences: their projected effects on crime and prison populations. *J. Crim. L. & Criminology*, *69*, 604.

Petitclerc, A., Gatti, U., Vitaro, F., & Tremblay, R. E. (2013). Effects of juvenile court exposure on crime in young adulthood. *Journal of child psychology and psychiatry*, *54*(3), 291-297.

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.

Phelps, M. S. (2011). Rehabilitation in the punitive era: The gap between rhetoric and reality in US prison programs. *Law & society review*, *45*(1), 33-68.

Poe-Yamagata, E. (2009). *And justice for some: Differential treatment of minority youth in the justice system*. DIANE Publishing.

Poulson, B. (2003). A third voice: A review of empirical research on the psychological outcomes of restorative justice. *Utah L. Rev.*, 167.

**Povitsky 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.

Pranis, K. (1997). From vision to action: Some principles of restorative justice. *Church and Society*, *87*(4), 32-42.

Pratt, T. C. (2018). *Addicted to incarceration: Corrections policy and the politics of misinformation in the United States*. Sage Publications.

Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology*, *38*(3), 931-964.

Pratt, T. C., & Godsey, T. W. (2003). Social support, inequality, and homicide: A cross-national test of an integrated theoretical model. *Criminology*, *41*(3), 611-644.

Presser, L., & Van Voorhis, P. (2002). Values and evaluation: Assessing processes and outcomes of restorative justice programs. *Crime & Delinquency*, *48*(1), 162-188.

**Quay, H. C., & Love, C. T. (1977). The effect of a juvenile diversion program on rearrests. *Correctional Psychologist, 4*(4), 377-396.

**Quinn, W. H., & Van Dyke, D. J. (2004). A multiple family group intervention for first-time juvenile offenders: Comparisons with probation and dropouts on recidivism. *Journal of Community Psychology, 32*(2), 177-200.

Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2008). Does unconscious racial bias affect trial judges. *Notre Dame L. Rev.*, *84*, 1195.

*Rausch, S. (1983). Court processing versus diversion of status offenders: A test of deterrence and labeling theories. *Journal of Research in Crime and Delinquency, 20*(1), 39-54.

Redlich, A. D., Bushway, S. D., & Norris, R. J. (2016). Plea decision-making by attorneys and judges. *Journal of Experimental Criminology, 12*(4), 537-561.

Reid, M. M. (2016). The Culture of Mass Incarceration: Why" locking them up and throwing away the key" Isn't Working and How Prison Conditions Can Be Improved. U. Md. LJ *Race Relig., Gender & Class*.

Reiss, A. J., & Tonry, M. H. (1986). *Communities and crime* (Vol. 8). University of Chicago Press Journals.

Richardson, L. S. (2016). Systemic Triage: Implicit Racial Bias in the Criminal Courtroom. *Yale LJ, 126*, 862.

*Riestenberg, N. (4 March 2014). Restorative group conferencing and sexting: Repairing harm in Wright County. Cyberbullying Research Center. http://cyberbullying.us/restorative-groupconferencing- and-sexting/.

Roach, K. (2000). Changing punishment at the turn of the century: Restorative justice on the rise. *Canadian J. Criminology*, *42*, 249.

*Roberts, L. (1998). Victim Offender Mediation: An Evaluation of the Pima County Juvenile Court Center's Victim Offender Mediation Program (VOMP). Available from Frasier Area Community Justice Initiatives in Langley, British Columbia.

Roberts, D. E. (2003). The social and moral cost of mass incarceration in African American communities. *Stan. L. Rev.*, *56*, 1271.

Rodriguez, N. (2007). Restorative justice at work: Examining the impact of restorative justice resolutions on juvenile recidivism. Crime & Delinquency, 53(3), 355-379.

*Roseman, C. P., Ritchie, M., & Laux, J. M. (2009). A restorative justice approach to empathy development in sex offenders: An exploratory study. *Journal of Addictions & Offender Counseling, 29*(2), 96-109.

*Roy, S. (1993). Two types of juvenile restitution programs in two Midwestern counties: A comparative study. *Federal Probation, 57*(4), 48-53.

*Ruback, R. B., Cares, A. C., & Hoskins, S. N. (2006). Evaluation of best practices in restitution and victim compensation orders and payments. Harrisburg, PA: Pennsylvania Commission on Crime and Delinquency.

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American journal of sociology*, *94*(4), 774-802.

Santos, C. M. D. C., Pimenta, C. A. D. M., & Nobre, M. R. C. (2007). The PICO strategy for the research question construction and evidence search. *Revista latino-americana de enfermagem*, *15*(3), 508-511.

*Schafer, N. E. (1988). Evaluation of the Alaska Pre-Trial Intervention Program.

Schanzenbach, M. M., & Tiller, E. H. (2008). Reviewing the sentencing guidelines: Judicial politics, empirical evidence, and reform. *The University of Chicago Law Review*, *75*(2), 715-760.

Schiavone, S. K., & Jeglic, E. L. (2009). Public perception of sex offender social policies and the impact on sex offenders. *International Journal of Offender Therapy and Comparative Criminology*, *53*(6), 679-695.

Schiff, M. (1998). Restorative justice interventions for juvenile offenders: A research agenda for the next decade. *Western Criminology Review*, *1*(1), 1-16.

*Schiff, M., & Bazemore, G. (2012). Whose kids are these? Juvenile justice and education partnerships using restorative justice to end the school-to-prison pipeline. Keeping Kids in School and Out of Courts, 68.

Schmitt, J., Warner, K., & Gupta, S. (2010). The high budgetary cost of incarceration. *Washington, DC: Center for Economic and Policy Research*.

Schneider, A. (1985). Guide to juvenile restitution programs. *Washington, DC: National Criminal Justice Reference Service and GPO*.

Schur, E. M. (1971). *Labeling deviant behavior*. New York: Harper & Row.

Seidman, I., & Pokorak, J. (2011). Justice responses to sexual violence. In M. P. Koss, J. W. White, & A. Kazdin (Eds.), *Violence against women and girls, Vol. 2. Navigating the solutions* (pp. 137-158). Washington, DC: American Psychological Association.

Seidman, I., & Vickers, S. (2005). The second wave: An agenda for the next thirty years of rape law reform. *Suffolk University Law Review*, *38*, 457-490.

**Severy, L. J., & Whitaker, J. M. (1982). Juvenile diversion: An experimental analysis of effectiveness. *Evaluation Review*, *6*(6), 753-774.

*Shaff, C., Bright, S., & Cavanagh, T. (2007). Restorative Justice at Colorado State. *Journal of Student Affairs*, 16, 10.

Shapiro, C. (1990). Is restitution legislation the chameleon of the victims' movement? In B. Galaway & J. Hudson (Eds.), *Criminal justice, restitution, and reconciliation* (p. 73-SQ). Monsey, NY: Willow Tree.

Shaw, C. R., & McKay, H. D. (1942). Juvenile delinquency and urban areas.

**Shelden, R. G. (1999). Detention Diversion Advocacy: An Evaluation. *Juvenile Justice Bulletin*.

Sherman, L., Neyroud, P. W., & Neyroud, E. (2016). The Cambridge Crime Harm Index: measuring total harm from crime based on sentencing guidelines. Policing: A Journal of Policy and Practice, 10(3), 171-183.

Sherman, L. W., Strang, H., Mayo-Wilson, E., Woods, D. J., & Ariel, B. (2015). Are restorative justice conferences effective in reducing repeat offending? Findings

from a Campbell systematic review. *Journal of Quantitative Criminology*, *31*(1), 1-24.

**Shichor, D., & Binder, A. (1982). Community restitution for juveniles: An approach and preliminary evaluation. *Criminal Justice Review, 7*(2), 46-50.

*Shichor, D., & Sechrest, D. K. (1998). A comparison of mediated and non-mediated juvenile offender cases in California. *Juvenile and Family Court Journal*, *49*(2), 27-40.

*Shichor, D., Sechrest, D. K., & Matthew, R. (2000). Victim-Offender Mediation in Orange County, California. Institute for Conflict Management, St. Vincent de Paul Center for Community Reconciliation (Santa Ana, Calif., February 2000)

*Shinar, D., & Compton, R. P. (1995). Victim impact panels: Their impact on DWI recidivism.

Simon, J. (2014). *Mass incarceration on trial: A remarkable court decision and the future of prisons in America*. New Press, The.

Singer, R. G. (1970). Prison Conditions: An Unconstitutional Roadblock to Rehabilitation. *Cath. UL Rev., 20*, 365.

Sloane, D. C., & Choi, H. (2016). Crime and Community Well-Being: The Role of Social Capital and Collective Efficacy in Increasing Safety. In Social Factors and Community Well-Being (pp. 87-99). Springer, Cham.

Sloan, J. J., Smykla, J. O., & Rush, J. P. (2004). Do juvenile drug courts reduce recidivism?: Outcomes of drug court and an adolescent substance abuse program. *American Journal of Criminal Justice*, *29*(1), 95-115.

Spamann, H., & Klöhn, L. (2016). Justice is less blind, and less legalistic, than we thought: Evidence from an experiment with real judges. *The Journal of Legal Studies*, *45*(2), 255-280.

Spohn, C. (2000). Thirty years of sentencing reform: The quest for a racially neutral sentencing process. *Criminal justice, 3*, 427-501.

Spohn, C., & Holleran, D. (2002). The effect of imprisonment on recidivism rates of felony offenders: A focus on drug offenders. *Criminology*, *40*(2), 329-358.

Spohn, C., Piper, R. K., Martin, T., & Frenzel, E. D. (2001). Drug courts and recidivism: The results of an evaluation using two comparison groups and multiple indicators of recidivism. *Journal of drug issues*, *31*(1), 149-176.

Spohn, C., & Spears, J. (1996). The effect of offender and victim characteristics on sexual assault case processing decisions. *Justice Quarterly*, *13*(4), 649-679.

Steffensmeier, D., & Demuth, S. (2000). Ethnicity and sentencing outcomes in US federal courts: Who is punished more harshly?. *American sociological review*, 705-729.

Steffensmeier, D., Ulmer, J., & Kramer, J. (1998). The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology*, *36*(4), 763-798.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of clinical epidemiology*, *54*(10), 1046-1055.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomized controlled trials. *Bmj*, *343*, d4002.

*Stewart, M. J. (2008). An outcomes study of juvenile diversion programs on non-serious delinquent and status offenders (Doctoral dissertation, Case Western Reserve University).

*Stinchcomb, J. B., Bazemore, G., & Riestenberg, N. (2006). Beyond zero tolerance: Restoring justice in secondary schools. *Youth Violence and Juvenile Justice*, *4*(2), 123-147.

*Stone, K. J. (2000). An evaluation of recidivism rates for resolutions northwest's victim-offender mediation program.

*Strode, E. (1997). Victims of Property Crime Meeting Their Juvenile Offenders: Victim Participants' Evaluation of the Dakota County (MN) Community Corrections Victim Offender Meeting Program: a Project Based Upon a Co-investigation with the Center for Restorative Justice and Mediation, School of Social Work, University of Minnesota, St. Paul, MN (Doctoral dissertation, Smith College School for Social Work).

Sullivan, D., & Tifft, L. (2008). *Handbook of restorative justice: A global perspective*. Routledge.

*Szmania, S. J., & Mangis, D. E. (2005). Finding the right time and place: A case study comparison of the expression of offender remorse in traditional justice and restorative justice contexts. *MArq. l. reV.*, 89, 335.

Tangney, J. P., Stuewig, J., & Hafez, L. (2011). Shame, guilt, and remorse: Implications for offender populations. *Journal of Forensic Psychiatry & Psychology*, *22*(5), 706-723.

The Sentencing Project. (2018, June). *Facts Sheet: Trends in U.S. Corrections*. Retrieved from http://www.sentencingproject.org

Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted?. *Statistics in medicine*, *21*(11), 1559-1573.

*Torbet, P., Ricci, R., Brooks, C., & Zawacki, S. (2001). Evaluation of Pennsylvania's school-based probation program. Pittsburgh, PA.: National Center for Juvenile Justice. Retrieved December, 2, 2013.

Travis, J. (2014). Assessing the State of Mass Incarceration: Tipping Point or the New Normal?. *Criminology & Public Policy*, *13*(4), 567-577.

Tyler, T. R. (1984). The Role of Perceived Injustice in Defendant's Evaulations of Their Courtroom Experience. *LAw & Soc'y REv.*, *18*, 51.

Ulmer, J., & Bradley, M. (2006). Variation in trial penalties among serious violent offenses. *Criminology, 44*, 631–670.

*Umbreit, M. S. (1988). Mediation of victim offender conflict. *J. Disp. Resol.*, 85.

Umbreit, M. S. (1989). Crime victims seeking fairness, not revenge: Toward restorative justice. *Fed. Probation*, 53, 52.

*Umbreit, M. S. (1991). Minnesota Mediation Center produces positive results. *Corrections Today, 53*(5), 192-196.

*Umbreit, M. S. (1993). Juvenile offenders meet their victims: The impact of mediation in Albuquerque, New Mexico. Family Court Review, 31(1), 90-100.

**Umbreit, M. S. (1994a). Crime victims confront their offenders: The impact of a Minneapolis mediation program. *Research on Social Work Practice*, 4, 436-447.

Umbreit, M. S. (1994b). *Victim meets offender: The impact of restorative justice and mediation*. Monsey, NY: Criminal Justice Press.

Umbreit, M. S. (2002). *The handbook of victim offender mediation: An essential guide to practice and research*. John Wiley & Sons.

Umbreit, M. S. (1995). Holding juvenile offenders accountable: A restorative justice perspective. *Juvenile and Family Court Journal*, *46*(2), 31-42.

Umbreit, M. S. (1997). Humanistic mediation: A transformative journey of peacemaking. *Mediation Quarterly*, *14*(3), 201-213.

Umbreit, M. S. & Armour, M. P. (2010). *Restorative justice dialogue: An essential guide for research and practice*. New York, NY: Springer.

**Umbreit, M. S., & Coates, R. B. (1992). The impact of mediating victim offender conflict: An analysis of programs in three states. *Juvenile and Family Court Journal, 43*(1), 21-28.

Umbreit, M. S., Coates, R. B., & Kalanj, B. (1994). *Victim meets offender: The impact of restorative justice and mediation* (pp. 53-64). Monsey, NY: Criminal Justice Press.

Umbreit, M. S., Coates, R. B., & Vos, B. (2002). *The impact of restorative justice conferencing: A review of 63 empirical studies in 5 countries.* University of Minnesota Center for Restorative Justice & Peacemaking, School of Social Work, University of Minnesota.

Umbreit, M. S., Coates, R. B., & Vos, B. (2007). Restorative justice dialogue: A multi-dimensional, evidence-based practice theory. *Contemporary Justice Review, 10*(1), 23-41.

*Umbreit, M. S., & Greenwood, J. (1999). National survey of victim-offender mediation programs in the United States. *Mediation Quarterly,* 16(3), 235-251.

*Umbreit, M., Lewis, T., & Burns, H. (2003). A community response to a 9/11 hate crime: Restorative justice through dialogue. *Contemporary Justice Review, 6*(4), 383-391.

*Umbreit, M. S., & Vos, B. (2000). Homicide survivors meet the offender prior to execution: Restorative justice through dialogue. *Homicide Studies, 4*(1), 63-87.

*Umbreit, M. S., Vos, B., Coates, R. B., & Brown, K. (2003). Victim offender dialogue in violent cases: The Texas and Ohio experience. *VOMA Connections, 14*(1), 12-17.

United States Courts. (2014). *Fiscal Year Funding and Cost Containment Initiatives*. Retrieved from http://www.uscourts.gov

United States Courts. (2018). *Annual Report, Funding and Budget.* Retrieved from http://www.uscourts.gov

*Urban, L. S., & Burge, S. E. (2006). Victim/offender mediation in St. Louis: An assessment. *Justice Research and Policy, 8*(2), 89-114.

Van Ness, D. W. (1989). Pursuing a restorative vision of justice. *Justice: The restorative vision,* 11-27.

Van Ness, D., Carlson, D., Crawford, T., & Strong, R. (1989). Restorative justice practice. *Monograph. Washington, DC: Justice Fellowship*.

Van Ness, D. & Heetderks, K. (1997). Restoring justice. *Cincinnati: Anderson Publishing Company*.

Van Wormer, K. (2009). Restorative justice as social justice for victims of gendered violence: A standpoint feminist perspective. *Social Work, 54*(2), 107-116.

Wagner, P. (2003). The Prison Policy Initiative. *The Prison Index: Taking the Pulse of the Crime Control Industry.* Retrieved from http:prisonpolicy.org

Walgrave, L., & Bazemore, G. (1999). Restorative juvenile justice: In search of fundamentals and an outline for systemic reform.

Walker, M. U. (2006). Restorative justice and reparations. *Journal of Social Philosophy, 37*(3), 377-395.

Walker, L. (2013). Restorative justice: definition and purpose. *Restorative justice today: practical applications,* 3-13.

Walker, J. T. (2007). Eliminate residency restrictions for sex offenders. *Criminology & Pub. Pol'y*, *6*, 863.

*Walker, L., & Greening, R. (2010). Huikahi restorative circles: A public health approach for reentry planning. *Fed. Probation*, 74, 43.

*Walker, L., & Hayashi, L. A. (2007). Pono Kaulike: A Hawaii criminal court provides restorative justice practices for healing relationships. *Fed. Probation*, 71, 18.

Warner, T. D., & Kramer, J. H. (2009). Closing the revolving door? substance abuse treatment as an alternative to traditional sentencing for drug-dependent offenders. *Criminal Justice and Behavior*, *36*(1), 89-109.

Watson, R., Stimpson, A., & Hostick, T. (2004). Prison health care: a review of the literature. *International journal of nursing studies*, *41*(2), 119-128.

*Wax, M. L. (1977). The effects of symbolic restitution and presence of victim on delinquent shoplifters (Doctoral dissertation, Washington State University).

Wehrman, M. M. (2010). Race, concentrated disadvantage, and recidivism: A test of interaction effects. *Journal of Criminal Justice*, *38*(4), 538-544.

Weitekamp, E. G., & Kerner, H. J. (Eds.). (2012). Restorative justice: theoretical foundations. Routledge.

Weitekamp, E. G., & Parmentier, S. (2016). Restorative justice as healing justice: looking back to the future of the concept.

Western, B., & Pettit, B. (2010). Incarceration & social inequality. *Daedalus*, *139*(3), 8-19.

Wheeldon, J. (2009). Finding common ground: restorative justice and its theoretical construction (s). *Contemporary Justice Review*, *12*(1), 91-100.

*White, C. A. C. (2000). The promise of restorative justice: An outcomes evaluation of an Orange County Victim Offender Reconciliation Program, with focus on the victim's perspective.

**Wiinamaki, L. A. (1999). Victim-offender reconciliation programs: Juvenile property offender recidivism and severity of reoffense in three Tennessee counties.

Wildra, E. (2017, December 7). Prison Policy Initiative. *Tracking the Impact of the Prison System on the Economy.* Retrieved from http://www.prisonpolicy.org

Wildeman, C., & Western, B. (2010). Incarceration in fragile families. *The future of children*, *20*(2), 157-177.

Willis, G. W., & Grace, R. C. (2009). Assessment of community reintegration planning for sex offenders: Poor planning predicts recidivism. *Criminal Justice and Behavior, 36*, 494- 512.

Wilson, D. [The Campbell Collaboration]. (2011, September 24). *Effect Size Calculation and Basic Meta-Analysis, David B. Wilson* [YouTube]. Retrieved from http:// www.youtube.com/watch?reload=9&v=nkcZFAmmKeE.

Wilson, D. [The Campbell Collaboration]. (2013, May 29). *Calculating Effect Sizes, David Wilson* [YouTube]. Retrieved from http://www.youtube.com/watch?v=Fggs7zOhw6c.

Wilson, R. J., McWhinnie, A., Picheca, J. E., Prinzo, M., & Cortoni, F. (2007). Circles of support and accountability: Engaging community volunteers in the management of high-Risk sexual offenders. *The Howard Journal of Criminal Justice*, *46*(1), 1-15.

Wilson, D. B., Olaghere, A., Kimbrell, C. S., George Mason University, & United States of America. (2017). Effectiveness of Restorative Justice Principles in Juvenile Justice: A Meta-Analysis.

Winnick, T. A., & Bodkin, M. (2008). Anticipated stigma and stigma management among those to be labeled “ex-con”. *Deviant Behavior*, *29*(4), 295-333.

Wozniak, J. F., Braswell, M. C., & Vogel, R. E. (Eds.). (2008). Transformative justice: critical and peacemaking themes influenced by Richard Quinney. Lexington Books.

*Wynne, J., & Brown, I. (1998). Can Mediation Reduce Offending?. *Probation journal*, 45, 21-26.

Yazzie, R. (1994). Life comes from it: Navajo justice concepts. *NML Rev.*, *24*, 175.

Yoon, S. (2015). Why Do Victims Not Report?: The Influence of Police and Criminal Justice Cynicism on the Dark Figure of Crime.

Zehr, H. (1990). Changing lenses: A new focus for crime and justice. Scottsdale, PA: Herald Press.

Zehr, H. (2011). Restorative or transformative justice. Restorative justice blog.

Zehr, H. (2015). The little book of restorative justice: revised and updated. Skyhorse Publishing, Inc.

Zehr, H. & Mika, H. (1998). Fundamental concepts of restorative justice. *Contemporary Justice Review*, 1, 47–55.

Zehr, H., & Umbreit, M. (1982). Victim offender reconciliation: An incarceration substitute. *Fed. Probation*, *46*, 63.

APPENDIX A

CODING MANUAL

1. Publication status: 0=no; 1=yes

2. Publication type: 0=thesis or dissertation; 1=book or book chapter; 2=peer-reviewed journal article or manuscript; 3=technical report

3. Diversion used: 0=no; 1=yes; 99=not specified (n/s)

4. Diversion delivery: 0=pre-adjudication; 1=post adjudication; 2=other; 3=mixed; 98=not applicable (n/a); 99 n/s

5. Program type circle: 0=no; 1=yes; 99=n/s

6. Program type mediation? 0=no; 1=yes; 99=n/s

7. Program type conferencing? 0=no; 1=yes; 99=n/s

8. Do victim and offender meet to dialogue (family as victim, surrogate victims included)? 0=no; 1=yes; 98=n/a; 99=n/s

9. Was any member of the community involved (facilitator or support members not included)? 0=no; 1=yes; 98=n/a; 99=n/s

10. Program type restitution? 0=no; 1=yes; 99=n/s

11. Restitution type: 0=community service, 1=monetary compensation, 2=personal; 3=work agreement; 4=mixed; 98=n/a; 99=n/s

12. Personal restitution agreement: (text)

13. Program type other than circle, conference, mediation, restitution: 0=no; 1=yes; 98=n/a; 99=n/s

14. Program type other: (text)

15. Program type hybrid model (encompasses multiple restorative elements or types combined as one delivery package): 0=no; 1=yes; 98=n/a; 99=n/s

16. More than one program type compared separately: 0=no; 1=yes; 98=n/a; 99=n/s

17. Program title(s): (text); 99=n/s

18. State(s) or region(s): (text); 99=n/s

19. Offender age group: 0=juveniles only; 1=adults only; 2=mixed; 98=n/a; 99=n/s

20. Context of facilitation: 0=schools; 1=criminal justice (diversion, court, corrections); 2=community; 3=other; 4=mixed; 98=n/a; 99=n/s

21. Facilitation other: (text)

22. Referral type: 0=schools; 1=probation/parole; 2= court; 3=law enforcement; 4=community; 5=corrections (re-entry); 6=other; 7=mixed; 98=n/a; 99=n/s

23. Referral type other: (text)

24. Offender criminal history: 0=no; 1=yes; 2=mixed; 98=n/a; 99=n/s

25. Offense type category: 0=sex crime; 1=non-violent; 2=violent; 3=mixture; 98=n/a; 99=n/s

26. Offense type (list all separated by commas): 1=arson; 2=assault and battery; 3=auto theft; 4=burglary; 5=child abuse/neglect; 6=criminal mischief; 7=criminal misconduct; 8=criminal trespassing; 9=domestic violence/intimate partner violence; 10=drug crime; 11=drunk driving/ impaired driving offenses; 12=financial/white collar; 13=harassment; 14=kidnapping; 15=larceny; 16=rape/statutory rape; 17=robbery; 18=school violations; 19=sexual assault; 20=shoplifting; 21=simple assault; 22=stalking; 23=juvenile status offenses (e.g., curfew violations, underage drinking); 24=theft; 25=threats; 26=vandalism; 27=mixed; 28=other (text); 98=n/a; 99=n/s

27. Offense type other: text; 98=n/a

28. Methodological Strength Score (raw):
    1=Either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. No use of control variables in statistical analysis to adjust for differences between treated and untreated groups or periods;
    2=Use of adequate control variables and either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. In (a), control variables or matching techniques used to account for cross-sectional differences between treated and controls groups. In (b), control variables are used to account for before-and-after changes in macro level factors;
    3=Comparison of outcomes in treated group after an intervention, with outcomes in the treated group before the intervention, and a comparison group used to provide a counterfactual (e.g. difference in difference). Justification given to choose of comparator group that is argued to be similar to the treatment group. Evidence presented on comparability of treatment and control groups. Techniques such as regression and (propensity score) matching may be used to adjust for difference between treated and untreated groups, but there are likely to be important unobserved differences remaining; 4=Quasi-randomness in treatment is

exploited, so that it can be credibly held that treatment and control groups differ only in their exposure to the random allocation of treatment. This often entails the use of an instrument or discontinuity in treatment, the suitability of which should be adequately demonstrated and defended;
5= Reserved for research designs that involve explicit randomization into treatment and control groups, with Randomized Control Trials (RCTs) providing the definitive example. Extensive evidence provided on comparability of treatment and control groups, showing no significant differences in terms of levels or trends. Control variables may be used to adjust for treatment and control group differences, but this adjustment should not have a large impact on the main results. Attention paid to problems of selective attrition from randomly assigned groups, which is shown to be of negligible importance. There should be limited or, ideally, no occurrence of 'contamination' of the control group with the treatment.

29. Methodological Score Adjusted: 0=no; 1=1point adjustment; 2=2 point adjustment; 98=n/a; 99=n/s

30. Methodological Strength Score (adjusted): (text number 1-5); 98=n/a

31. SMS violation(s) type (if multiple separate by commas): 1=Randomization is successful; 2=Attrition carefully addressed or not an issue; 3=Contamination not an issue; 4=Instrument relevant (explains treatment); 5=Instrument exogenous (not explained by outcome); 6=Instrument excludable (does not directly affect outcome); 7=Discontinuity in treatment is sharp (e.g., strict eligibility requirement) or fuzzy discontinuity method used; 8=Only treatment changes at boundary; 9=Behavior is not manipulated to make the cut-off; 10=Key assumption of 'no anticipation' holds; 11=Variation in timing; 12=Selection equation includes relevant observable variables; 13=Control group would have followed same trend of treatment group; 14=Known time period for treatment; 15=fixed effect is at the unit of observation; 16=Year effects are included; 17=Appropriate time-varying controls are used; 18=Adequate control group established; 19=Treatment date is known and singular; 20=Selection equation includes relevant observable variables; 21=Good Matching variables (i.e., relevant to selection); 22=selection equation includes relevant observable variables; 23=Good Matching variables (i.e., relevant to selection; 24=Significant common support; 25=Adequate control variables are used; 98=n/a

32. Voluntary offender participation? 0=no (no admission of guilt or mandated participation?) 1=voluntary (condition of traditional processing sentence, meaning failure to participate would result back to mandate by to traditional system); 2=yes, (totally voluntary); 3=mixed; 98 n/a; 99=n/s

33. Longest follow-up (in months); 98 n/a; 99 n/s

34. Number of dependent variables: 1=1; 2=2 or more (multiple)

35. Recidivism: 1=not specified; 2=subsequent disciplinary referral; 3=self-reported delinquency; 4=subsequent court referrals; 5=parole/probation violations; 6=any future contact with criminal justice system; 7=number of re-offenses (arrests or convictions); 8=re-arrest in general; 9=re-arrest specific crime type; 10=number days until re-arrest; 11=reconviction; 12=re-incarceration; 13=risk of recidivism; 14=re-offense severity: 98=n/a

36. Relations/cohesion: 0=peer relations/cohesion 1=family relations/cohesion; 2=community relations/cohesion; 98=n/a

37. Academic achievement: 0=no; 1=yes

38. Peer aggression: 0=no; 1=yes

39. Offender social competence: 0=no; 1=yes

40. Satisfaction: 0=participant satisfaction (general); 1=victim satisfaction; 2=offender satisfaction; 3=citizen/community satisfaction; 4=legal agent satisfaction: 98=n/a

41. Victim fear of revictimization: 0=no; 1=yes

42. Victim post-traumatic stress symptoms: 0=no; 1=yes

43. Victim anxiety: 0=no; 1=yes

44. Victim anger: 0=no; 1=yes

45. Fairness: 0=participant perception of fairness (general); 1=victim perception of fairness; 2=offender perception of fairness; 3=community/citizen perception of fairness; 98=n/a

46. Restitution compliance: 0= restitution compliance (general); 1=monetary restitution compliance; 2=work agreement restitution compliance; 3=community service restitution compliance; 98=n/a

47. Stakeholder obligation agreement reached: 0=no; 1=yes

48. Offender program completion rate: 0=no; 1=yes

49. Sense of community involvement: 0=no; 1=yes

50. Stakeholder perceptions of the value in process: 0=no; 1=yes

51. Measure of reparation of harms: 0=no; 1=yes

52. Restoration of relationships: 0=no; 1=yes

53. Measures of offender accountability in general: 0=no; 1=yes

54. Apology written or verbal: 0=no; 1=yes

55. Offender expressed remorse: 0=no; 1=yes

56. Offender expressed guilt: 0=no; 1=yes

57. Offender empathy: 0=no; 1=yes

58. Offender shame: 0=no; 1=yes

59. Offender parole success scale: 0=no; 1=yes

60. Offender length of employment time after reentry in months: 0=no; 1=yes

61. Offender expressions of honesty in self-reported behavior: 0=no; 1=yes

62. Offender expression of responsibility or accountability: 0=no; 1=yes

63. School attendance after treatment: 0=no; 1=yes

64. Cost comparison between treatment(s) and control: 0=gross/net in dollars; 1=per individual case in dollar

65. Victim mental health scores: 0=no; 1=yes

66. Victim sense of hope: 0=no; 1=yes

67. Additional coder notes: (text)

APPENDIX B

SMS SCORE ADJUSTMENT GUIDELINES

(Madaleno & Waights, 2010, p. 37-33)

Method	Maximum SMS score (method, implementation)	Adjusted SMS score (method, implementation)
Randomised Control Trial (RCT) ***a.k.a.*** **Field Experiment**	5, 5 if • Randomisation is successful • Attrition carefully addressed or not an issue • Contamination not an issue	5, 4 if • One of the criteria is severely violated 5,3 if • Two or more of the criteria are severely violated
Instrumental Variable (IV) ***a.k.a.*** **Two-Stage Least Squares (2SLS)**	4, 4 if instrument • Relevant (explains treatment) • Exogenous (not explained by outcome) • Excludable (does not directly affect outcome)	Scored as per underlying method if • Instrument invalid e.g cross section with invalid IV scores 2; difference-in-difference with invalid IV scores 3 (see below)
Regression Discontinuity Design (RDD)	4, 4 if • Discontinuity in treatment is sharp (e.g. strict eligibility requirement) or fuzzy discontinuity method used • Only treatment changes at boundary • Behaviour is not manipulated to make the cut-off	Scored as per underlying method if • Discontinuity conditions severely violated e.g cross section with invalid IV scores 2; difference-in-difference with invalid IV scores 3 (see below)

Method		Maximum SMS score (method, implementation)	Adjusted SMS score (method, implementation)
Difference in differences (DID) **a.k.a.** **Dif in dif**		3,3 if • Control group would have followed same trend and treatment group • Known time period for treatment	3, 2 if • Either of the criteria is not satisfied
Panel methods	**Panel Fixed Effects (FE)**	3, 3 if • Fixed effect is at the unit of observation • Year effects are included • Appropriate time-varying controls are used	3, 2 if • One or more of the three criteria is not satisfied
	First Differences (FD)	3, 3 if • Year effects are included • Appropriate time-varying controls are used	3, 2 if • Either of the two criteria is not satisfied
	Arellano-Bond (AB)	3, 3 if • Year effects are included • Appropriate time-varying controls are used	3, 2 if • One of the two criteria is not satisfied

Method		Maximum SMS score (method, implementation)	Adjusted SMS score (method, implementation)
Hazard Regressions	**Mixed Proportional Hazards (MPH)**	4, 4 if • Key assumption of 'no anticipation' holds • Variation in timing (e.g. people start training at different times relative to becoming unemployed)	4, 3 if • Anticipation of treatment likely • Little variation in timing 4, 2 if • Neither of the criteria is satisfied
	Proportional Hazards (PH)	3,3 if • Adequate control group is established • Treatment date is known and singular	3,2 if • One of the two criteria is not satisfied

Method		Maximum SMS score (method, implementation)	Adjusted SMS score (method, implementation)
Heckman Two-Stage Approach (H2S) **Or** **Control Function (CF)**	**With IV**	4,4 if • Selection equation includes an IV that satisfies the three criteria (see IV)	Scored as per underlying method if • Instrument invalid e.g cross section with invalid IV scores 2; difference-in-difference with invalid IV scores 3 (see below)
	With DID or panel method	3,3 if • Selection equation includes relevant observable variables • DID or panel criteria met	3, 2 if • One of the criteria is not satisfied
	Cross-sectional	2, 2 if • Selection equation includes relevant observable variables	2, 1 if • Selection equation does not include relevant observable variables
Propensity Score Matching (PSM) ***a.k.a.*** **Matching**	**With DID or panel method**	3, 3 if • Matching criteria satisfied (see below) • DID or panel criteria satisfied	3, 2 if • One of the criteria is violated
	Cross-sectional	2, 2 if • Good matching variables (i.e. relevant to selection) • Significant common support	2, 1 if • One of the criteria is violated

Method	Maximum SMS score (method, implementation)	Adjusted SMS score (method, implementation)
Cross-sectional regression	2, 2 if • Adequate control variables are used	2, 1 if • Inadequate control variables
Before-and-after	2, 2 if • Adequate control variables are used	2, 1 if • Inadequate control variables
Stated effects/impact ***a.k.a.*** **Additionality** **Stated experiences**	1, 1 if • Fact that observed differences in outcomes are not necessarily effects of the policy is acknowledged • No control group	

APPRENDIX C

SMS SCORES TABLE

Study ID	Final SMS Score *adjusted scores
*Acosta, Chinman, Ebener, Phillips, Xenakis, & Malone, 2016	4
Armour, 2012	1
*Baffour, 2006	3
*Baglivio & Jackowski, 2013	4
Baker, 2008	2
Beck-Zierdt, 1980	1
*Behtz, 2004	4
Berger, Lipsey, Dennison, & Lange, 1977	3
Bergseth & Bouffard, 2007	4
*Bergseth & Bouffard, 2012	3
Bouffar, Cooper, & Bergseth, 2017	3
Bradshaw & Umbreit, 1998	1
Brooks, 2013	4
*Butts & Snyder, 1992	3
Butts, Buck, & Coggeshall, 2002	3
*Cannon & Stanford, 1981	3
Carr & Nelson, 2000	3
Clarke, Valente, & Mace, 1992	3
Coates & Gehm, 1985	3

Coates, Umbreit, & Vos, 2000	1
Coates, Umbreit, & Vos, 2003	1
Coldren Jr, Haring, Luecke, Sintic, & Balgoyen, 2011	4
Cosden, Casas, & Wolfe, 1999	4
Crofoot, 1987	4
Crotty, J. and Meier, 1980	3
*Davis, 2009	4
de Beus & Rodriguez, 2007	4
DeWitt & DeWitt, 2012	1
Dick, 1999	3
Duwe, 2018	5
*Ezell, 1986	2
Featherston, 2014	5
Fercello & Umbreit, 1999	2
Fishbein, Davis, & Hamparin, 1984	1
Flaten, 1996	2
Forgays & DeMilio, 2005	3
Fors & Rojek, 1999	4
Fox, 2013	1
Freeman, 2018	1
Fulkerson, 2001	5
Galaway, 1998	2
Gase, Kuo, Lai, Stoll, & Ponce, 2016	5

*Griffith, 1983	3
Guedalia, 1979	1
Haarman & Covington, 1981	1
Harrison, Maupin, & Mays, 2001	1
Heinz, Galaway, & Hudson, 1976	3
Helfgott, Lovell, Lawrence, & Parsonage, 2000	2
Henggeler, Melton, & Smith, 1992	5
Hitao, 1999	3
Hughes & Schneider, 1989	1
Hughes & Schneider, 1990	1
Jackson, 2009	2
*Jeong, McGarrell, & Kroovand Hipple, 2012	4
Karp & Drakulich, 2004	1
Koch, 1986	2
Koss, 2014	2
Lane, Turner, Fain, & Sehgal, 2007b	5
Lane, Turner, Fain, & Sehgal, 2005	5
Lee, 1999	3
Leslie, 2002	1
Levi, 1982	1
Lewis, 2009	1
Lincoln, Teilmann, Klein, & Labin, 1977	5
*Lipsey, Cordray, & Berger, 1981	4

McCold, 1998	3
*McGarrell & Kroovand Hipple, 2007	4
McGarrell, Olivares, Crawford, & Kroovand, 2000	5
McMorris, Beckman, Shea, Baumgartner, & Eggert, 2013	1
*Mills, Barocas, & Ariel, 2012	4
Minor, Wells, Soderstrom, Bingham, & Williamson, 1999	3
Myers, Burton, Sanders, Donat, Cheney, Fitzpatrick, & Monaco, 2000	3
Nelson, 2000	2
*Newland, 1980	3
*Niemeyer & Shichor, 1996	2
Norris, 2008	1
Northcutt Bohmert, Duwe, & Kroovand Hipple, 2016	1
*Norton, Gold, & Peralta, 2013	2
Nugent & Paddock, 1995	3
Nugent & Paddock, 1996	2
Patrick, Marsh, Bundy, Mimura, & Perkins, 2004	5
Povitsky Stickle, Connell, Wilson, & Gottfredson, 2008	5
Quay & Love, 1997	4
Quinn & Van Dyke, 2004	3
Rausch, 1983	4
Riestenberg, 2014	1
Roberts, 1998	1

*Roseman, Ritchie, & Laux, 2009	2
Roy, 1993	2
Ruback, Cares, & Hoskins, 2006	2
Schafer, 1988	1
Schiff & Bazemore, 2012	3
*Severy & Whitaker, 1982	4
Shaff, Bright, & Cavanagh, 2007	2
Shelden, 1999	3
Shichor & Binder, 1982	4
Shichor & Sechrest, 1998	1
Shichor, Sechrest, & Matthew, 2000	1
Shinar & Compton, 1995	3
Stewart, 2008	2
Stinchcomb, Bazemore, & Riestenberg, 2006	1
Stone, 2000	3
Strode, 1997	1
Szmania & Mangis, 2005	1
Torbet, Ricci, Brooks, & Zawacki, 2001	1
Umbreit & Coates, 1992	3
Umbreit & Greewood, 1999	2
Umbreit & Vos, 2000	1
Umbreit, 1988	1
Umbreit, 1991	1

Umbreit, 1993	2
Umbreit, 1994a	4
Umbreit, Lewis, & Burns, 2003	1
Umbreit, Vos, Coates, & Brown, 2003	1
Urban & Burge, 2006	3
Walker & Greening, 2010	2
Walker & Hayashi, 2007	1
*Wax, 1977	4
White, 2000	1
*Wiinamaki, 1997	4
Wynne, 1998	2