

## Why do evaluation researchers in crime and justice choose non-experimental methods?<sup>1</sup>

CYNTHIA LUM\*.\*.\*

*College of Criminal Justice, Northeastern University, Boston, MA 02115, USA*

*\*corresponding author: E-mail: c.lum@neu.edu*

SUE-MING YANG

*Department of Criminology, University of Maryland, College Park, MD 20742, USA*

**Abstract.** Despite the general theoretical support for the value and use of randomized controlled experiments in determining ‘what works’ in criminal justice interventions, they are infrequently used in practice. Reasons often given for their rare use include that experiments present practical difficulties and ethical challenges or tend to over-simplify complex social processes. However, there may be other reasons why experiments are not chosen when studying criminal justice-related programs. This study reports the findings of a survey of criminal justice evaluation researchers as to their methodological choices for research studies they were involved in. The results suggest that traditional objections to experiments may not be as salient as initially believed and that funding agency pressure as well as academic mentorship may have important influences on the use of randomized controlled designs.

**Key words:** criminal justice evaluation, evaluation research, experiments, scientific validity, what works

No question has simultaneously dominated practice and research in criminology and criminal justice more so than ‘what works?’ in reducing crime, recidivism, and crime-related risk factors. Since a number of highly influential reports in the 1970s and 1980s indicating a grim future for many criminal justice programs and policies (see Kelling et al. 1974; Lipton et al. 1975; Spelman and Brown 1984), the push towards evaluating criminal justice interventions to find effective treatments has defined the role of a number of researchers. This emphasis on program effectiveness can be seen in systematic reviews of programs (most notably, Sherman et al.’s 1997 report to congress), the increased use of meta-analyses to draw more parsimonious conclusions from the plethora of evaluation research (see, e.g., Andrews et al. 1990; Cox et al. 1995; Dowden et al. 2003; Lipsey and Wilson 1993; Logan and Gae 1993; Lösel and Kofler 1989; Prendergast et al. 2000; Wilson 2000, 2001; Wilson et al. 2000, 2001; Whitehead and Lab 1989), and the establishment of the Campbell Collaboration,<sup>2</sup> an organization which advocates for higher quality research and evidence-based policy (see Farrington and Petrosino 2001; Petrosino et al. 2001).

---

\*.\*In August 2005, Dr. Lum’s affiliation will change to George Mason University.

A natural development from this 'what works' pursuit has become assessing the quality of these evaluations. The believability of evaluation research depends not only on the theoretical sense of what is being evaluated but also upon the evaluation's methodological quality. For example, we are cautious of the results of studies which evaluate the effectiveness of drug treatment or incarceration using a sample of individuals who most likely would not re-offend even without any intervention. Ensuring that scientifically valid approaches are used when evaluating the effects of treatment is imperative when asserting that a treatment or policy 'works' or 'doesn't work' in reducing crime, criminality, or crime-related risk factors (Cook and Campbell 1979; Farrington 2003a; Farrington and Petrosino 2001; Shadish et al. 2002; Sherman et al. 1997; Weisburd and Petrosino, forthcoming).

Broadly, scientific validity emphasizes that the methodology used in an evaluation of a criminal justice intervention maintains certain standards that contribute to greater believability in asserted conclusions. Although different types of scientific validity have been articulated (see Cook and Campbell 1979; Farrington 2003a), scholars have argued that internal and external validity are especially important to methodological quality (Farrington 2003a; Farrington and Petrosino 2001; Shadish et al. 2002). Specifically, external validity refers to "the generalizability of causal relationships across different persons, places, times, and operational definitions of interventions and outcomes" (Farrington 2003a: 54). External validity can be maximized by choosing random samples from a population (Farrington 2003a), replicating the treatment on different samples and conditions, and continually evaluating the intervention (McCord 2003). Internal validity refers to an evaluator's ability to determine whether the intervention did in fact cause a change in the outcome measured or that treatment effects can be clearly distinguished from other effects (Shadish et al. 2002: 97). Internal validity is often maximized through the use of the experimental design as the evaluation methodology.

An experimental design establishes internal validity by randomly allocating a population of interest (or sample thereof) into different conditions, treatments, or programs to isolate the effects of those conditions from other possible factors that may contribute to group differences. Random allocation of treatment programs ensures that there is no systematic bias that divides subjects into treatment and control groups (Campbell and Stanley 1963; Farrington and Petrosino 2001). Specifically, random allocation allows for the assumption of equivalence between treatment and comparison groups, a necessary condition to 'rule out' other confounding factors that might explain differences between groups after treatment (Weisburd 2003). Thus, as Cook (2003) emphasizes, random allocation provides an appropriate counterfactual in the control group, showing what would happen had the treatment not been administered. Therefore, when carefully designed and implemented, a randomized controlled experiment is regarded as highly useful in contributing to the believability of the results of evaluation research (Boruch et al. 2000a; Burtless 1995; Cook 2003; Sherman 2003; Weisburd 2000, 2001).

The use of experiments has been supported not only on these scientific and statistical grounds in determining 'what works,' but also, as Cook (2003) points out,

empirical evidence indicates that real differences can exist between results of experiments and non-experiments. For criminal justice experiments, Weisburd et al. (2001) found that non-experimental evaluations in criminal justice tended to result in more positive or 'it works' findings compared to experimental evaluations, perhaps leading to false conclusions about program effectiveness (see Gordon and Morse, 1975, who found similar findings in social research generally). Furthermore, meta-analysts have found differences in the size or magnitude of effects depending on the evaluation method used. Effect sizes in experimental evaluations can be larger (see Wilson et al. 2001), smaller (see Wilson et al. 2000), or without significant difference (see Lipsey and Wilson 1993; Whitehead and Lab 1989) compared to non-experiments. Some have also justified the importance of experimental over non-experimental methods on other grounds, including that it is unethical to *not* use randomized experiments to discover whether a program is effective or harmful (Boruch 1976; McCord 2003; Weisburd 2003) or that experimentation benefits policy and normative practice (Boruch et al. 2000a; Cook 2003). Clearly there is, at least in theory, justification for the use of randomized experiments in evaluating the effects of social programs.

### **The choice to use experiments in criminal justice evaluations**

Despite this general methodological justification for the use of the randomized controlled experiment, this type of design is infrequently used when evaluating criminal justice interventions (Shepherd 2003). In Sherman et al. (2002), the most comprehensive collection to date of criminal justice evaluations in the United States, of the 657 evaluations listed and summarized, 84% used non-experimental methods to draw conclusions about treatments while only 16% used an experimental methodology. A variety of reasons have been hypothesized and well documented that might account for this large discrepancy.

The most common arguments against the use of randomized experiments often involve practical or ethical concerns (for a review of some of these arguments, see Boruch 1976; Clarke and Cornish 1972; Cook 2003; Farrington 1983; Shepherd 2003; Stufflebeam 2001; Weisburd 2000). Experimentation is seen as difficult to conduct in non-clinical settings either due to a variety of problems such as implementation issues (Boruch 1976; Petersilia 1989), convincing practitioners to participate (Feder et al. 2000), or ethical or moral dilemmas in treating some individuals and not others based on a random allocation scheme (Boruch et al. 2000b; Clarke and Cornish 1972). Many of these arguments do not challenge experimentation in theory, but rather recognize the limitations of randomized controlled experiments in practice. Others have also argued that the use of experimental designs may be inadequate in capturing the complex social or research environment, and some have challenged its limited use or its ability to maximize methodological quality (see Burtless 1995; Clarke and Cornish 1972; Heckman and Smith 1995; Pawson and Tilley 1994, 1997; Stufflebeam 2001).

While reasons of practicality, ethics, and lack of complexity point specifically to concerns with the experimental methodology itself, there may also be other reasons besides methodological considerations that might influence researcher decisions. For example, early academic experiences, including the influence of mentors and academic advisors, may influence the methodological choices that a researcher makes in his or her academic career. Literature on academic socialization indicates that many of these factors can influence the success, productivity, and quality of a scientist's research (Corcoran and Clark 1984; Reskin 1979). Generally, academic mentorship is believed to be a positive influence on researcher productivity, job placement, and career development (Cameron and Blackburn 1981; Clark and Corcoran 1986; Raul and Peterson 1992). Along similar lines, mentorship might have important influences on the methodological choices a researcher makes. It may be the case that evaluation researchers who have worked with mentors or advisors on experiments in the past tend to go on to conduct experiments in the future.

The academic discipline from which researchers come or the formal academic training a researcher has received may also contribute to various methodological choices that evaluation researchers make when studying criminal justice interventions. Those conducting criminal justice evaluations come from a variety of disciplines and backgrounds; perhaps the shying away from experimentation comes from scientific biases within certain academic disciplines. Wanner et al. (1981) found differences across academic disciplines generally in terms of research productivity. Differences might also exist in theoretical and methodological norms as well as what is considered 'scientific' across different disciplines (Kuhn 1970). In terms of criminal justice evaluations, the most obvious differences might be reflected in biases toward certain subject matters. Psychologists may examine the effects of programs which attempt to affect risk factors, early childhood development, or psychological treatment in prisons, while those from the field of education may be more concerned about school-related programs. Those trained in criminology or criminal justice may focus on programs in traditional criminal justice institutions. Related to these choices might be differences in the disciplinary biases, nature, mechanisms, and subjects of research, which may also help to shape the choice of research method. For example, the reliance on experiments is much more common in psychology than sociology, criminology, or education because of the traditional use of laboratory experiments in psychological research. In some cases, as Palmer and Petrosino (2003) discovered, the replacing of psychologists with other social scientists in research agencies can also contribute to the decline in the use of experiments in evaluation research conducted by those agencies. The reliance on experiments is much more common in psychology than sociology, criminology, or education because of the traditional use of laboratory experiments in psychological research.

Studies have also indicated that government funding and external pressures can influence research methodology. In a historical account of the use of the randomized controlled experiment in criminal justice, Farrington (2003b) suggested that James Stewart of the National Institute of Justice (NIJ) was highly influential during the

1980s in advocating the use of experiments for projects receiving NIJ funding. Garner and Visher (2003), when reviewing the funding awarded by NIJ in the 1990s, found that the number of awards and the total amount of funds awarded to the research using randomized experiments declined across that time.<sup>3</sup> Palmer and Petrosino (2003) also have emphasized the importance of funding agencies on the choices of research methods when analyzing the changes and transition of the California Youth Authority (CYA). They found that when the National Institute of Mental Health (NIMH) was the main funding agency of CYA, the support of randomized trials by NIMH constituted a powerful incentive for the CYA to use them (Palmer and Petrosino 2003: 240). However, Palmer and Petrosino also discovered that when the Law Enforcement Administrator Authority (LEAA) became CYA's new funding agency, this changed the research orientation of CYA as LEAA encouraged the use of short-term, quick analysis over long-term experiments (Palmer and Petrosino 2003: 243–244).

While many of these arguments have been either hypothesized or made on a case-by-case basis, how salient are they? Cook (2003) recently raised this question for education researchers offering rebuttals to a wide variety of objections to experimental methods. We sought to further explore the pervasiveness of some of these traditional objections within the criminal justice field by understanding empirically the methodological choices of authors of criminal justice-related evaluations. In particular, does a researcher's affiliated discipline or academic training relate to his/her involvement with experiments as opposed to non-experiments? What are some of the reasons that researchers give for having chosen a methodological design for a specific study they were involved in, and what do they see were advantages or disadvantages in using either an experimental or non-experimental method? Are traditional objections as salient as proposed or are other factors at work? Finally, what are the general attitudes of researchers towards traditional objections to experiments, and do researchers of experiments and non-experiments differ in these attitudes? It is to these questions that we now turn.

### **The study**

To approach these questions, we decided to draw a sample from experimental and non-experimental criminal justice evaluations from which we could survey authors about their methodological choices. We chose to sample from evaluations rather than evaluators because we were interested in understanding methodological choices researchers made for specific studies of which the methodology was known as well as their general feelings about randomized experiments. Thus, we wished to draw a sample from the total population of studies which had evaluated the effects of treatment, interventions or programs related to crime, criminal justice, or crime-related risk factors. While we anticipated that researchers' responses about past research may be influenced by subsequent experiences or inaccurate due to the passing of time, we felt this was the most appropriate way of probing the aforementioned questions.

To simplify the identification of a population of criminal justice-related evaluations, we began with the most comprehensive collection currently available, Sherman et al.'s (2002) volume *Evidence-Based Crime Prevention*, also known as the updated "Maryland Report." In 1997, the Department of Criminology and Criminal Justice at the University of Maryland (College Park) was commissioned by the United States Congress to collect and evaluate all criminal justice interventions to determine, as the title refers to, "What Works, What Doesn't and What's Promising" in crime prevention (Sherman et al. 1997).<sup>4</sup> The collection of these studies included not only locating and compiling evaluations, but also determining what intervention, program, or treatment was being evaluated, what was the outcome of the evaluations, and what type of scientific method the author(s) used to determine their results.

The Maryland Report was a major undertaking involving six professors, eight scientific advisors, and over 20 graduate students, and the updated 2002 volume included further collaboration by other researchers. Evaluations were sought across multiple criminal justice and non-criminal justice institutions including programs in schools, families, communities, police, courts, corrections, labor markets, and places/situations. Undoubtedly, the Maryland Report and subsequently the updated 2002 volume did not include the entire universe of evaluation research. Evaluations included only those that were written or interpreted into the English language and that satisfied a threshold of methodological quality (Sherman et al. 1997).<sup>5</sup> However, it remains the most comprehensive collection currently available.

We also chose the updated Maryland Report to select our sample not only because of its comprehensiveness, but also because each study evaluated in the Maryland Report was evaluated for methodological rigor and type. The Maryland Report (and subsequently the updated report by Sherman et al. 2002 used here) systematically categorizes studies using a 'scientific methods scale' or the 'SMS,' which consists of scores '1,' '2,' '3,' '4' and '5' (Table 1). The SMS scores provided us with the ability to identify experiments and non-experiments in criminal justice evaluation studies. While the Maryland Report created the scientific method scale as an

Table 1. Sherman et al.'s Scientific Methods Scale (SMS).

<i>SMS score</i>	<i>Description</i>
1	Correlation between a crime prevention program and a measure of crime or crime risk factors
2	Temporal sequence between the program and the crime or risk outcome clearly observed, or a comparison group present without demonstrated comparability to the treatment group
3	A comparison between two or more units of analysis, one with and one without the program
4	Comparison between multiple units with and without the program, controlling for other factors, or a nonequivalent comparison group has only minor differences evident
5	Random assignment and analysis of comparable units to program and comparison groups

ordinal scale, we chose not to compare evaluations across the five categorizations, but rather divided the Maryland dataset into two groups – those which received a score of ‘5’ and those which did not. Specifically, studies with scores of ‘1’ through ‘4’ were categorized as ‘non-experiments’ while those with a score of ‘5’ were assigned to the ‘experiments’ category. The creation of this dichotomy was useful to explore our questions of interest because of the clearer methodological division between randomized and non-randomized studies and because we were interested in differences between these two groups. However, it should be noted that the Maryland Report authors suggest that individual studies could be downgraded on this scale (for example, from being initially considered a randomized experiment and then later to a non-randomized experiment) given certain rationales.<sup>6</sup>

Of the 657 studies listed in Sherman et al. (2002), 102 studies (16%) were given a score of ‘5’ and were deemed randomized controlled experiments while 555 (84%) were given scores of ‘1’ through ‘4’ and were classified as non-experiments. We began by randomly selecting 80 studies from the 102 randomized controlled experiments and 80 studies from the non-experiments for a total initial sample of 160 evaluation studies. We sought an equal sample from the experimental and non-experimental groups (therefore over-sampling from the randomized group) as our primary goal was to understand differences between these two groups. It should also be noted that because our unit of analysis was the evaluation and not the researcher, four studies chosen had matching authorships.

We then attempted to contact the first authors of each of the 160 sampled evaluations. We were interested in the first authors, anticipating that they were often the principal investigators of evaluation studies and would have more control, knowledge and understanding as to the decision making process in determining the methodology used to evaluate a particular program or treatment. If first authors could not be located or identified, we then sought responses from subsequent authors, if any. Of the 160 studies chosen from Sherman et al. (2002), authors for 131 studies could be located.<sup>7</sup> Of the studies in which authors were located, 66 were experiments and 65 were non-experiments.

Surveys were then sent to the authors of the 131 studies (see Appendix A). Each respondent was asked to complete the survey in reference to a study they conducted that was cited in Sherman et al. (2002). As Appendix A indicates, the survey asked a number of questions about the study authors including their current and graduation discipline, their current job position, their training in research methods, statistics and experimentation, as well as their general attitudes about randomized controlled experiments. For the specific study of interest, questions included why authors chose a particular methodology, whether they felt that choice was appropriate, and whether they had been influenced in making such a choice. They were also given the opportunity to list advantages and disadvantages in using the particular methodology that they chose. Ninety-three of the 131 study authors responded. Of these 93 responses, four respondents asserted that their studies were not evaluations<sup>8</sup> and six studies were excluded as they were later deemed to be duplicates of other sampled evaluations.

Thus, in the end, our response rate was 83 'valid' responses of 121 valid studies, or 70%. We had valid responses from authors of 46 experiments and 37 non-experiments.

## Findings

To explore the influence of academic discipline or methodological training on choices, we asked authors of each study in our sample about the current academic discipline with which they were associated (Table 2). Although differences may be due to chance because of small sample sizes, we found it interesting that criminologists were authors of a larger proportion of non-experimental evaluations (50%) more so than experimental ones (26%). However, authors who currently work in the fields of psychology or sociology represented greater proportions within the experimental group than compared to their non-experimental proportions. This trend was also evident when examining the discipline from which authors received their highest degree (94% had doctorates). Again, we found that criminologists tended to be over-represented in the non-experimental group than their counterparts in psychology or sociology (Table 3). And, we found that the year in which study authors' degrees were obtained<sup>9</sup> or the number of years between graduation and the study itself<sup>10</sup> were not significantly correlated to the author's choice of methodology. These preliminary results suggest that our initial hypothesis about disciplinary preferences may prove to be salient, especially differences between the fields of criminology and psychology.

The preference towards experimentation may not only be related to disciplinary norms but also to the methodological training of study authors. We found that the lack of formal training in research methods generally was not connected to a researcher's choice of methodology. The vast majority (93%) of authors of our sampled studies had taken research methods in graduate school, and most (88%) reported that experiments were discussed in these general courses. Nor did we find any statistically significant relationship between those who had used experimentation for the specified study of interest (or even among those who had ever been involved in a randomized controlled experiment) and whether they had taken a specific course in experimentation.

Table 2. Current academic discipline of study authors.

<i>Discipline</i>	<i>Non-experiments</i>	<i>Experiments</i>
Criminology/Criminal Justice	18 (50.0%)	12 (26.1%)
Psychology	7 (19.4%)	14 (30.4%)
Sociology	3 (8.3%)	5 (10.9%)
Others <sup>a</sup>	8 (22.3%)	15 (32.6%)
Total	36 (100.0%)	46 (100.0%)

<sup>a</sup>Specifically, public policy, education, child development, anthropology, social work, economics, public health, statistics and mathematics, medicine and communication studies.



Table 3. Discipline of highest degree obtained by study authors.

<i>Discipline</i>	<i>Non-experiments</i>	<i>Experiments</i>
Criminology/Criminal Justice	12 (32.4%)	7 (15.2%)
Psychology	12 (32.4%)	18 (39.1%)
Sociology	3 (8.1%)	12 (26.1%)
Others <sup>a</sup>	10 (27.0%)	9 (19.6%)
Total	37 (100%)	46 (100%)

<sup>a</sup>Specifically, public policy, education, child development, anthropology, social work, economics, public health, statistics and mathematics, medicine, public administration, political economics, and communication studies.

Although formal training was not necessarily related to whether or not individuals had chosen to conduct a randomized experiment, we acknowledged that often, graduate training in research (and further learning) does not result from formal classroom teaching but rather from mentorship experiences, interactions with colleagues, or self-learning. To gauge these influences, we asked authors of experiments to rank the sources of influence as to their knowledge of how to conduct randomized controlled experiments using a 5-point scale, 1 being the 'least influential' and 5 being the 'most influential.' Of the choices, education and colleagues were seen as influential (Figure 1). Furthermore, when comparing authors of experiments and non-experiments, those associated with experiments in our sample more often had students or colleagues go on to conduct other

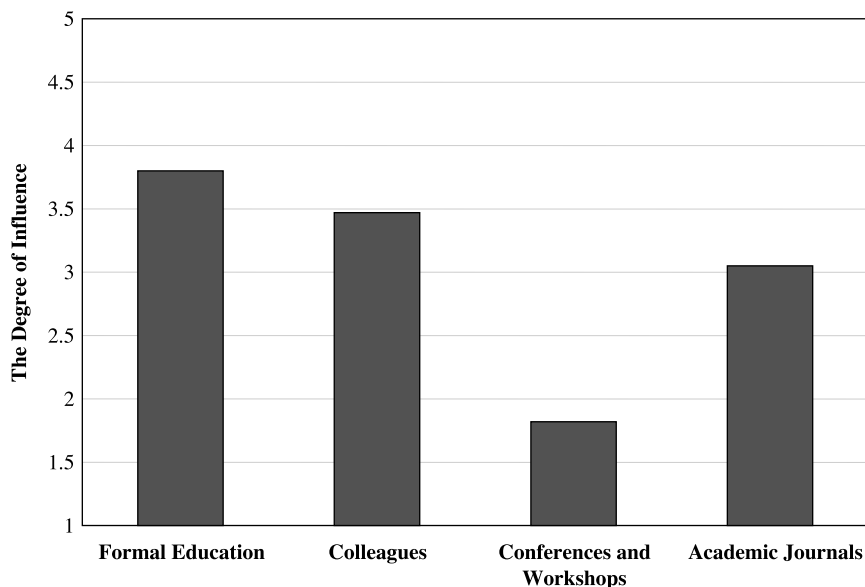


Figure 1. The sources of knowledge about how to conduct randomized experiments.

Table 4. Have students or colleagues trained with you on the specific project referred to gone on to conduct experimental designs on their own?

	<i>Non-experiments</i>	<i>Experiments</i>
Yes	11 (31.0%)	29 (66.0%)
No	23 (66.0%)	11 (25.0%)
Don't Know	1 (3.0%)	4 (9.0%)
Total	35 (100.0%)	44 (100.0%)

Chi-square = 13.282\*\* ( $p=0.001$ ).

randomized experiments compared with their non-experimental counterparts (Table 4), an association that was statistically significant. The findings suggest that perhaps experience, collegiality, and mentorship may matter as much as, if not more than, formal education in terms of whether researchers use experiments or non-experiments when conducting evaluation research for criminal justice evaluations.

We also sought to understand why researchers choose experimental or non-experimental designs in evaluating specific interventions. First, we asked study authors of both non-experiments and experiments whether they felt that the methodology they used was the most appropriate method in evaluating the specific treatment. Almost all of the researchers, regardless of the method they chose, considered their choices were the most appropriate at the time (Table 5), although a larger percentage of those conducting non-experiments answered 'no' compared to their experimental counterparts (this relationship was not statistically significant). It could be hypothesized that one more obvious reason why criminal justice evaluators choose non-experimental designs is because they believe these designs to be 'appropriate.'

To better understand why researchers of these evaluations felt that the methodology they used was 'appropriate,' we asked two related questions. The first was to probe whether individuals or agencies outside of the project had potentially influenced or pressured the individual into choosing a particular methodological design. We asked study authors in our sample whether an individual, institution, or group outside of themselves or their project staff recommended the methodological design used for the specific evaluation study. We gave respondents a number of

Table 5. Did the study's author feel that the methodology used was the 'most appropriate' method to evaluate the intervention?

	<i>Non-experiments</i>	<i>Experiments</i>
Yes	32 (87.0%)	44 (96.0%)
No	5 (14.0%)	2 (4.0%)
Total	36 (100.0%)	46 (100.0%)

Chi-square = 2.230 ( $p=0.137$ ).

sub-choices within the ‘yes’ choice as indicated in Table 6. When examining the differences between ‘yes’ and ‘no’ responses, those who conducted randomized controlled experiments were significantly more likely to have been influenced by an outside entity to use experimentation than those who had completed non-experimental evaluations. Furthermore, 26% of the sampled experimental studies’ authors reported that funding agency pressure to conduct experiments was an important reason why they chose an experimental method over a non-experimental one. These results indicate that outside influence, in particular funding agency pressures, may have significant effects in pushing individuals towards using experiments more often, confirming our initial hypothesis.

We also explored ‘appropriateness’ by asking study authors to cite advantages and disadvantages of the method they used for the specified evaluation sampled. While we acknowledge that authors may feel the methods they chose were appropriate even if citing disadvantages, this qualitative question helped us to probe authors for reasons behind their methodological choices. In general, experimenters felt more strongly about the advantages of experimentation than the disadvantages. The advantage of assuring causality was most cited, where authors of experiments said, for example, that experiments have the “ability to determine cause and effect,” can “accurately estimate program impact,” “allow us to draw strong conclusions,” that “strongest inferences about treatment effects can be made with randomized controlled trials,” or that experiments “control for threats to internal validity.” Reasons of efficiency were also noted, including that experiments were “the most powerful design with a small number of cases” or “the most efficient test of specific effects.” Furthermore, experimenters suggested that randomized controlled experiments tended to satisfy the requests of funding agencies or requirements of outside sources. Some disadvantages were cited as well, although not as often. Major disadvantages revolved around getting individuals or institutions to cooperate fully with the experiments as well as cost considerations.

For the non-experimenters, however, the focal concerns were somewhat different. Non-experimenters suggested that practicality was the biggest advantage

*Table 6.* Did an individual, institution or group outside of yourself and your project staff recommend the specific methodological design you used for this study?

	<i>Non-experiments</i>	<i>Experiments</i>
Yes	5 (13.9%)	16 (34.8%)
Funding agency requested	3 (8.0%)	12 (26.0%)
My employer requested	0 (0.0%)	2 (4.0%)
A colleague recommended	2 (5.0%)	2 (4.0%)
Another outside entity recommended	0 (0.0%)	0 (0.0%)
No	32 (87.0%)	30 (65.0%)
Total	37 (100.0%)	46 (100.0%)

Chi-square for ‘yes’ vs. ‘no’ = 4.908\* ( $p=0.023$ ).

with respect to the method they adopted. For example, some remarks included “advantages [of the specified non-experiment] lie in time and speed,” “other options were not practical,” “did not have funding or time for prospective approach,” or “flexibility of design.” Non-experimenters also felt non-experimental methods were advantageous when “no other control group was available” or because evaluations were “post hoc” and therefore “randomized experimentation was not possible.” Surprisingly, only one non-experimenter cited ethical concerns as the reason he/she did not use a randomized controlled experiment. Nonetheless, it is important to mention that a number of the non-experimenters thought that the main disadvantage of using non-experimental methods was that these designs could not control for the threat of internal validity, the main strength of experiments. For example, the “lack of controls,” “threats to internal validity,” “threats to interpretation of causal effects,” or that subjects were “not randomly assigned” were cited as disadvantages in using the design they chose.

While the above questions targeted specific studies, we also sought to determine study author’s general feelings about traditional objections related to experimentation. To gauge this, study authors were asked to rank four statements, “Randomized experimental design is the best method of linking cause and effect,” “Randomized experiments cannot be carried out ethically in criminal justice settings,” “Randomized experiments are often not practical in criminal justice research,” and “There are not enough randomized experiments in criminal justice research” using a five point scale of 1 = ‘strongly disagree,’ 2 = ‘disagree,’ 3 = ‘neither disagree nor agree,’ 4 = ‘agree,’ and 5 = ‘strongly agree.’

Before examining our results, we thought that perhaps representing study authors’ methodological preferences using one study may not accurately reflect whether or not they had ever conducted an experiment which may affect their general feelings about experimentation (and therefore cause problems when comparing the two groups which were based on the sampled studies). However, we found this not to be the case. Those who conducted non-experiments for the specified study sampled were significantly associated with the group membership of never having conducted an experiment (chi-square = 13.127,  $p < 0.001$ ). Furthermore, the number of randomized controlled experiments an author had ever conducted was positively related to whether or not they had, in the specified study sampled, used experimentation (logistic regression coefficient  $\beta = 0.191$ ,  $\exp(\beta) = 0.826$ ,  $p = 0.004$ ).

Interesting findings emerged when comparing the experimenters and non-experimenters. When asking study authors whether or not randomized experimentation was the best method of linking cause and effect (Figure 2), scholars of both experimental and non-experimental studies ranked this statement fairly high, suggesting that in general, criminal justice evaluators tend to believe that experimentation is the most appropriate method in determining causality. However, those who conducted experiments felt even more strongly about this statement than those who did not, a finding preliminarily reflected in our qualitative examination of the advantages and disadvantages authors gave for using the methodology in the

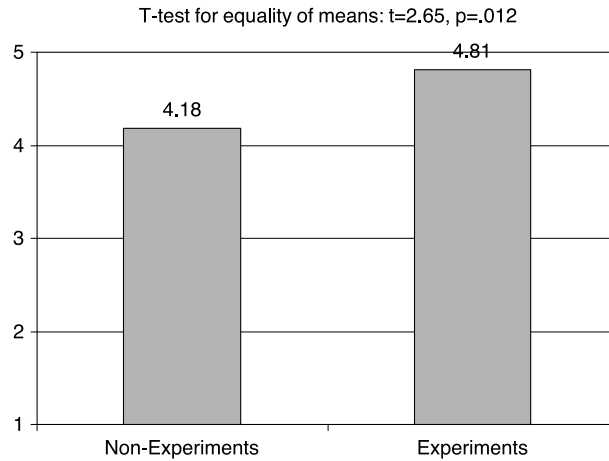


Figure 2. Randomized experimental design is the best method of linking cause and effect.

specific study of interest. The difference in the average rankings of this statement of those who had conducted an experiment for the specified study to those who had not was statistically significant.

Surprisingly, results showed that ethical concerns were not so much an issue as sometimes believed (see Weisburd 2000), which was also reflected in the qualitative analysis. As Figure 3 suggests, both experimental and non-experimental groups tended to disagree with the statement that experimentation cannot be carried out ethically in criminal justice settings. Although authors of experiments

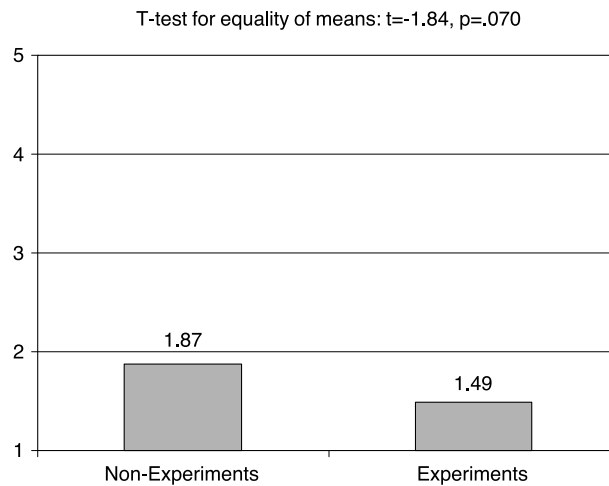


Figure 3. Randomized experiments cannot be carried out ethically in criminal justice settings.

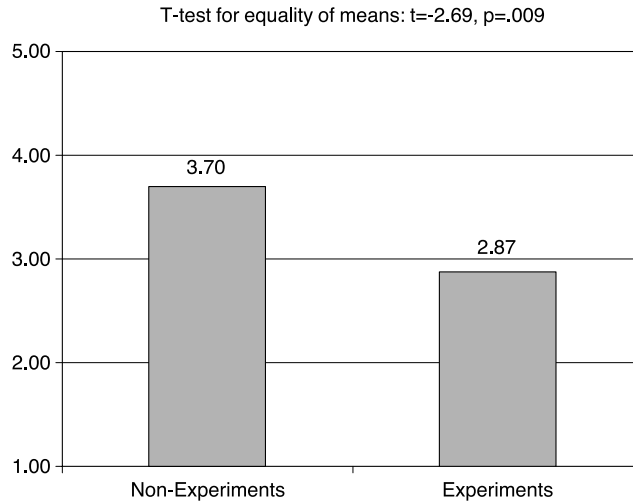


Figure 4. Randomized experiments are often not practical in criminal justice research.

on average more strongly disagreed, the difference between the two groups was not statistically significant.

However, issues of practicality were a concern for non-experimenters, again reflected in the qualitative assessment of advantages and disadvantages above. Figure 4 shows that authors of non-experiments were more likely to believe that experiments were not necessarily practical in criminal justice research, while those

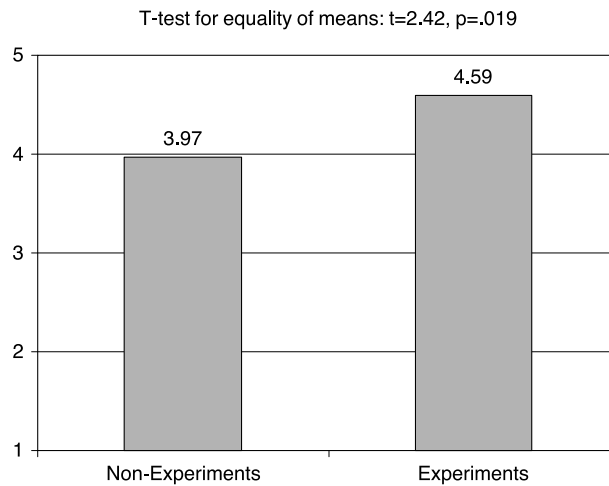


Figure 5. There are not enough randomized experiments in criminal justice research.

who had conducted experiments were more likely to disagree with this statement. However, when comparing Figures 2–4, it should be noted that while non-experimenters gave much higher scores to their worries of practicality than ethical concerns, nonetheless, they were still quite confident that experiments assured causality in criminal justice research. Thus, issues of practicality may be more of concern to all researchers than ethical problems, yet the belief that experimentation is the best way to assure causality remains salient.

Finally, when asking scholars whether or not there were ‘not enough’ randomized experiments in criminal justice research, both those involved with experiments and non-experiments tended to agree with this statement. When comparing those who had used experiments *versus* those who had not for the specific study sampled, this difference was statistically significant (Figure 5). Nonetheless, despite concerns of practicality, evaluation researchers generally tend to agree that more experimentation in criminal justice is warranted.

### **Discussion and conclusions**

These findings point to a number of interesting characteristics of evaluation researchers that may impact their methodological choices. In terms of training and disciplinary association, we discovered that the fields of criminology and criminal justice might be less encouraging of the utilization of randomized controlled experiments than psychology or sociology. While these nuances could be due to our sampling restrictions, they are nonetheless interesting given the recent findings by Palmer and Petrosino (2003). Criminology may be more policy-oriented, and therefore, may fall prey to short-term political goals that may discourage researchers from what they perceive as methodologies not conducive to public policy agendas. Differences between disciplinary norms may also be related to differences in research settings and units of analyses. For example, psychological experiments may be more frequently conducted in clinical or laboratory environments which might facilitate their use. Whatever the reason, as the study of criminology grows and as the term ‘criminologist’ more frequently means an individual who comes from the field of criminology, it may be useful to question and explore disciplinary norms if experiments are to be encouraged in criminal justice evaluations. No doubt, mechanisms such as the Campbell Collaboration, funding pressures, or the increased use of meta-analyses and systematic reviews can positively affect these norms.

Training is also an important aspect of disciplinary norms. Although formal education was generally considered important by the entire sample, we found that informal training mechanisms such as collegiality, mentorship, and experience may matter as much as formal education in terms of methodological decision making, a finding reflected in the literature on the sociology of higher education. Those who conducted experiments in criminal justice were more likely to have colleagues or students subsequently go on to conduct experiments than study authors of non-experiments (and those who conducted experiments chosen in our sample were

more likely to conduct more experiments over their academic lifespan). Furthermore, colleagues seemed to matter as much as formal education in terms of the sources of knowledge on how to conduct experiments. This reinforces what is often generally the case in both graduate and professional training; often methods of research and other skills are learned through research experiences and mentorship, rather than exclusively from formal classroom settings. The importance of mentorship and academic collegiality may therefore not only have its known benefits, but also can increase the strength of methodological quality of evaluation research and affect the scientific believability of academic research in general.

We also discovered that while collegiality may be one reason why experiments are perpetuated, funding agencies and other outside pressures were also very relevant. Specifically, those who conducted an experiment were more likely to have been influenced by an outside entity such as funding agencies in terms of choosing experimentation. The findings reinforce Farrington's findings cited at the beginning of this study of the influence of the National Institute of Justice on the use of experiments in the 1980s. Thus, an important policy implication of these findings is that requirements set by such agencies and funding pressures can influence the methods that scientists use. In particular, the power of the purse can encourage the use of more rigorous designs, which are often connected to problems of implementation and practicality.

Finally, we discovered surprising findings with regards to traditional objections to experimentation as well as general feelings about experimental and non-experimental methods. The advantage of assurance of causality remains a strong factor in not only study authors choosing experimentation but also when examining the general beliefs of all study authors about experiments. Practicality was considered a major concern of non-experimenters in our study and often influenced their methodological choices. However, these non-experimenters generally acknowledge that a major disadvantage of the methods they used was that these methods cannot assure internal validity to the extent that experimentation can; a finding confirmed by high ratings of the statement "Randomized experimental design is the best method of linking cause and effect." And, practicality, while trumping ethical concerns, did not surpass feelings of the causal worthiness of experimentation. Furthermore, other arguments against randomized controlled experiments were not present in the responses of authors, such as arguments challenging the rationale behind evaluation generally (see, e.g., Pawson and Tilley 1994, 1997) or statistical challenges to experimental methods themselves (see Heckman and Smith 1995). Rather, there was general consensus that experimentation was the best approach available, but that other practical reasons impeded researchers from using this methodology.

In total, our study suggests that traditional objections may not be as salient as perhaps initially believed, at least not to our sampled evaluators. If experimentation is positively viewed as enhancing the quality of evaluation research in criminal justice, it is fairly safe to say that there is much hope in promoting the use of experiments in evaluating the effects of treatment. This can be



accomplished through addressing and overcoming practical concerns, urging funding agencies to require more rigorous evaluation designs, and encouraging disciplinary norms (particularly within the discipline of criminology) towards experimentation. Clearly, the traditional objections to experimentation are not insurmountable.

### **Acknowledgements**

The authors would like to thank David Weisburd of the University of Maryland and Hebrew University who suggested the initial idea for this paper as well as Anthony Petrosino, David Farrington, and the anonymous reviewers for their thoughtful comments. Most importantly, the authors wish to acknowledge and thank the participants of our survey for their time and cooperation.

### **Appendix A: Survey instrument**

#### *General questions*

1. Current academic discipline with which you are associated (i.e., “Criminology,” “Sociology,” “Education,” etc.):
2. Current (or retired from) professional position (“Professor,” “Researcher,” “Practitioner,” “Government consultant,” etc.):
3. Highest academic degree awarded
  - (a) Bachelors
  - (b) Masters
  - (c) Professional Degree (law, medicine, business, etc.)
  - (d) Doctorate
4. Discipline in which highest academic degree was awarded:
5. Year highest degree was awarded:
6. In the course of your graduate studies, did you take a course on advanced statistical methods?
  - (a) yes
  - (b) no
7. In the course of your graduate studies, did you take a course specifically on randomized experimental designs?
  - (a) yes
  - (b) no
8. In the course of your graduate studies, did you take a course on general research methods?
  - (a) yes
  - (b) no

- 8a. Were randomized experimental designs discussed in this course?
- (a) yes
  - (b) no

*Regarding your specific study*

This section specifically references the study of which you were the primary author entitled “<<INSERT CITATION HERE>>.”

1. According to your recollection, what was the methodological design you used for this study to evaluate the criminal justice program or treatment?
  - (a) non-experimental method as described above
  - (b) quasi-experimental method as described above
  - (c) randomized experimental method as described above
  - (d) other, please specify here:
2. Did an individual, institution or group outside of yourself and your project staff recommend the specific methodological design you used for this study?
  - (a) Yes, the funding agency or their program managers for this project requested this specific methodology
  - (b) Yes, my employer (i.e., university, private research organization, governmental research group) requested that I use this specific methodology
  - (c) Yes, a colleague recommended that we use this particular methodology
  - (d) Yes, another outside entity recommended we use this methodology
  - (e) No, no agent outside of myself or my project staff recommended or required the use of this methodology.
3. Did you consider the methodology you used the most appropriate methodology for examining this research problem?
  - (a) yes
  - (b) no [Go to 3a.]
- 3a. If you did not consider the methodology you used the most appropriate, what methodology did you feel was more appropriate?
  - (a) non-experimental method
  - (b) quasi-experimental method
  - (c) randomized experimental method
4. For this specific study, what were the advantages and disadvantages in using the chosen methodology?

*Experience with randomized experiments*

1. Have you ever conducted or participated in a research program involving random allocation of subjects or other units to treatment and control conditions?
  - (a) Yes
  - (b) No (please skip to question 4)

2. If yes, in how many randomized controlled experiments have you been involved?
3. Of those mentioned in #2, for how many have you been a principal investigator?
4. On a scale from 1 to 5, 1 being the least influential and 5 being the most influential, please score the influence that each of the following sources of training or knowledge had in your learning about how to conduct a randomized experimental study.
  - (a) Formal education (i.e., college, graduate school)
  - (b) Influence of a colleague who was an expert in this area
  - (c) Conference, workshop or class
  - (d) Academic journal articles or books
  - (e) Other sources, please specify here:
6. Have students or colleagues trained with you on the specific project referred to above gone on to conduct experimental designs on their own?
  - (a) Yes
  - (b) No
7. On a scale from 1 to 5, 5 being that you highly agree and 1 being that you highly disagree, how do you feel about the following statements?
  - (a) All else being equal, a randomized experimental design is the best method of linking cause and effect.
  - (b) Randomized experiments cannot be carried out ethically in criminal justice settings
  - (c) Randomized experiments are often not practical in criminal justice research
  - (d) There are not enough randomized experiments in criminal justice research

## Notes

- 1 This paper was developed for and presented at the Third Annual Jerry Lee Crime Prevention Symposium, held on February 23, 2004 at the University of Maryland, College Park and was also presented at the Societies of Criminology 1st Key Issues Conference in Paris on May 15, 2004.
- 2 Modeled after the Cochrane Collaboration, the Campbell Collaboration (see [www.campbellcollaboration.org](http://www.campbellcollaboration.org)) is an international organization which prepares rigorous and systematic reviews of evaluation research in the social sciences and updates them on a periodic basis (see Farrington and Petrosino 2001, for a more complete description). The Crime and Justice Coordinating Group (<http://www.aic.gov.au/campbellcj/>) focuses specifically on criminal justice related programs.
- 3 The percentage of the experimental studies funded by NIJ ranged from 0% to 5.3%. In 1997, 1999, and 2000, there was no funding awarded to randomized experimental studies.
- 4 The 1997 University of Maryland report to Congress can be obtained at [www.ncjrs.org/works](http://www.ncjrs.org/works).
- 5 For a complete explanation of the methodology used in the report please see Chapter 2 and the Appendix of the Maryland Report located at <http://www.ncjrs.org/works/chapter2.htm> and <http://www.ncjrs.org/works/appendix.htm>, respectively.
- 6 MacKenzie (2002) explains that downgrading from, for example, a SMS score of “5” (a randomized controlled experiment) to a SMS score of “4” (a “quasi-experiment”)

- occurred when the random assignment was not successful, there were too few subjects, or when the attrition rates were high (MacKenzie 2002: 334). In these cases, the randomization scheme was seen to have lost its benefits of avoiding systematic bias.
- 7 Some authors of studies were deceased or retired and could not be located.
  - 8 There were some studies in the Maryland Report and the updated Sherman et al. (2002) volume that did not directly or explicitly involve crime, recidivism, criminality, or related risk factors that were initially captured in our sample but were later excluded. Furthermore, as Lawrence Sherman pointed out in personal correspondence with primary author, one study we sampled was identified as a review of others' research and therefore authors were not the original authors of the study. This study was therefore excluded from consideration.
  - 9 Logistic regression coefficient  $\beta = -0.010$ ,  $\exp(\beta) = 0.990$ ,  $p = 0.703$ .
  - 10 Logistic regression coefficient  $\beta = -0.016$ ,  $\exp(\beta) = 1.016$ ,  $p = 0.592$ .

## References

- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P. & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology* 28(3), 369–404.
- Boruch, R. (1976). On common contentions about randomized field experiments. In G. Glass (Ed.), *Evaluation studies review annual*. Beverly Hills, CA: Sage Publications.
- Boruch, R., Snyder, B. & DeMoya, D. (2000a). The importance of randomized field trials. *Crime and Delinquency* 46(2), 156–180.
- Boruch, R., Victor, T. & Cecil, J. S. (2000b). Resolving ethical and legal problems in randomized experiments. *Crime and Delinquency* 46(3), 330–353.
- Burtless, G. (1995). The case for randomized field trials in economic and policy research. *The Journal of Economic Perspectives* 9(2), 63–84.
- Cameron, S. & Blackburn, R. (1981). Sponsorship and academic career success. *The Journal of Higher Education* 52, 369–377.
- Campbell, D. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, American Educational Research Association.
- Clark, S. & Corcoran, M. (1986). Perspectives on the professional socialization of women faculty: A case of accumulated disadvantage? *The Journal of Higher Education* 57, 20–43.
- Clarke, R. V. & Cornish, D. B. (1972). The controlled trial in institutional research: Paradigm or pitfall for penal evaluators? *Home Office Research Studies* (Vol. 15). London, UK: Her Majesty's Stationery Office.
- Cook, T. (2003). Resistance to experiments: Why have educational evaluators chosen not to do randomized experiments? *The Annals of the American Academy of Political and Social Science* 589, 114–149.
- Cook, T. & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Corcoran, M. & Clark, S. (1984). Professional socialization and contemporary career attitudes of three faculty generations. *Research in Higher Education* 20, 131–153.
- Cox, S. M., Davidson, W. S., & Bynum, T. S. (1995). A meta-analytic assessment of

- delinquency-related outcomes of alternative education programs. *Crime and Delinquency* 2, 219–234.
- Dowden, C., Antonowicz, D. & Andrews, D. A. (2003). Effectiveness of relapse prevention with offenders: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology* 47(5), 516–528.
- Farrington, D. (1983). Randomized experiments on crime and justice. In M. Tonry & N. Morris (Eds.), *Crime & justice: An annual review of research* (Vol. IV, pp. 257–308). Chicago, IL: The University of Chicago Press.
- Farrington, D. (2003a). Methodological quality standards for evaluation research. *Annals of the American Academy of Political and Social Sciences* 587, 49–68.
- Farrington, D. (2003b). A short history of randomized experiments in criminology: A meager feast. *Evaluation Review* 27(3), 218–227.
- Farrington, D. & Petrosino, A. (2001). The Campbell Collaboration Crime and Justice Group. *Annals of the American Academy of Political and Social Sciences* 578, 35–49.
- Feder, L., Jolin, A. & Feyerherm, W. (2000). Lessons from two randomized experiments in criminal justice settings. *Crime and Delinquency* 46(3), 380–400.
- Garner, J. H. & Visser, C. A. (2003). The production of criminological experiments. *Evaluation Review* 27(3), 316–335.
- Gordon, G. & Morse, E. V. (1975). Evaluation research: A critical review. *The Annual Review of Sociology*.
- Heckman, J. & Smith, J. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives* 9(2), 85–110.
- Kelling, G., Pate, A. M., Dieckman, D. & Brown, C. E. (1974). *The Kansas City preventive patrol experiment: Summary report*. Washington DC: The Police Foundation.
- Kuhn, T. (1970). *The structure of scientific revolutions*. 2nd edn. Chicago, IL: University of Chicago Press.
- Lipsey, M. & Wilson, D. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist* 48, 1181–1209.
- Lipton, D., Martinson, R. & Wilks, J. (1975). *The effectiveness of correctional treatment: A survey of treatment evaluation studies*. New York: Praeger.
- Logan, C. H. & Gaes, G. G. (1993). Meta-analysis and the rehabilitation of punishment. *Justice Quarterly* 10, 245–263.
- Lösel, F. & Kofler, P. (1989). Evaluation research on correctional treatment in West Germany: A Meta-analysis. In H. Wegener, F. Lösel & J. Haisch (Eds.), *Criminal behavior and the justice system: Psychological perspectives*. New York: Springer-Verlag.
- MacKenzie, D. L. (2002). Reducing the criminal activities of known offenders and delinquents: Crime prevention in the courts and corrections. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie (Eds.), *Evidence based crime prevention* (pp. 330–404). London, UK: Routledge.
- McCord, J. (2003). Cures that harm: Unanticipated outcomes of crime prevention programs. *Annals of the American Academy of Political and Social Sciences* 587, 16–30.
- Palmer, T. & Petrosino, A. (2003). The “experimenting agency”: The California Youth Authority Research Division. *Evaluation Review* 27(3), 228–266.
- Pawson, R. & Tilley, N. (1994). What works in evaluation research? *British Journal of Criminology* 34(3), 291–306.
- Pawson, R. & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Petersilia, J. (1989). Implementing randomized experiments – lessons from BJA’s intensive supervision project. *Evaluation Review* 13(5), 435–458.
- Petrosino, A., Boruch, R., Soydan, H., Duggan, L. & Sanchez-Meca, J. (2001). Meeting the

- challenges of evidence-based policy: The Campbell Collaboration. *Annals of the American Academy of Political and Social Sciences* 578, 14–34.
- Prendergast, M. L., Podus, D., & Chang, E. (2000). Program factors and treatment outcomes in drug dependence treatment: An examination using meta-analysis. *Substance Use and Misuse* 35(12–14), 1931–1965.
- Raul, S. & Peterson, L. (1992). Nursing education administrators: Level of career development and mentoring. *Journal of Professional Nursing* 8, 161–169.
- Reskin, B. (1979). Academic sponsorship and scientists' careers. *Sociology of Education* 52, 129–146.
- Shadish, W., Cook, T. & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inferences*. Boston: Houghton-Mifflin.
- Shepherd, J. P. (2003). Explaining feast of famine in randomized field trials. *Evaluation Review* 27(3), 290–315.
- Sherman, L. W. (2003). Misleading evidence and evidence-led policy: Making social science more experimental. *The Annals of the American Academy of Political and Social Science* 589, 6–19.
- Sherman, L. W., Gottfredson, D., MacKenzie, D. L., Eck, J., Reuter, P. & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising: A report to the United States Congress*. Washington, DC: National Institute of Justice.
- Sherman, L. W., Farrington, D. P., Welsh, B. C. & MacKenzie, D. L. (Eds.), (2002). *Evidence based crime prevention*. London, UK: Routledge.
- Spelman, W. & Brown, D. K. (1984). *Calling the police: Citizen reporting of serious crime*. Washington, DC: United States Government Printing Office.
- Stufflebeam, D. L. (2001). *Evaluation models*. San Francisco, CA: Jossey-Bass.
- Wanner, R., Lewis, L. & Gregorio, D. (1981). Research productivity in academia: A comparative study of the sciences, social sciences and humanities. *Sociology of Education* 54, 238–253.
- Weisburd, D. (2000). Randomized experiments in criminal justice policy: Prospects and problems. *Crime and Delinquency* 46(2), 181–193.
- Weisburd, D. (2001). Magic and science in multivariate sentencing models: Reflections on the limits of statistical methods. *Israel Law Review* 35(2), 225–248.
- Weisburd, D. (2003). Ethical practice and evaluation of interventions in crime and justice: The moral imperative for randomized trials. *Evaluation Review* 27(3), 336–354.
- Weisburd, D. & Petrosino, A. (forthcoming). Experiments: Criminology. In K. Kempf (Ed.), *Encyclopedia of social measurement*. Chicago, IL: Academic Press.
- Weisburd, D., Lum, C. & Petrosino, A. (2001). Does research design affect study outcomes? *The Annals of the American Academy of Political and Social Science* 578, 50–70.
- Wilson, D. (2000). Meta-analyses in alcohol and other drug abuse treatment research. *Addiction* 95(3), 419–438.
- Wilson, D. (2001). Meta-analytic methods for criminology. *Annals of the American Academy of Political and Social Sciences* 578, 71–89.
- Wilson, D., Gallagher, C. & MacKenzie, D. L. (2000). A meta-analysis of corrections-based education, vocation, and work programs for adult offenders. *Journal of Research in Crime and Delinquency* 37, 347–368.
- Wilson, D., Gottfredson, D. & Najaka, S. (2001). School-based prevention of problem behaviors: A meta-analysis. *Journal of Quantitative Criminology* 17(3), 247–272.
- Whitehead, J. & Lab, S. (1989). A meta-analysis of juvenile correctional treatment. *Journal of Research in Crime and Delinquency* 26(3), 276–295.

**About the authors**

**Cynthia Lum** is an Assistant Professor in the College of Criminal Justice at Northeastern University and specializes in policing, place-based criminology and evaluation research. She has published a series of papers regarding evaluation research methodology in the *Annals of the American Academy of Political and Social Sciences* and is now conducting a Campbell systematic review of counter-terrorism strategies.

**Sue-Ming Yang** is a Doctoral candidate in the Department of Criminology and Criminal Justice at the University of Maryland. Her current research interests include criminological theory testing, etiology of violent behavior, advanced applied statistics and understanding the relationships between crime and place over time.