REDUCING PRISON MISCONDUCTS

What Works!

SHEILA A. FRENCH
PAUL GENDREAU
University of New Brunswick

A meta-analysis was conducted to assess the effectiveness of correctional treatment for reducing institutional misconducts. Sixty-eight studies generated 104 effect sizes involving 21,467 offenders. Behavioral treatment programs produced the strongest effects ($r = .26$, CI = .18 to .34). The numbers of criminogenic needs targeted and program therapeutic integrity were found to be important moderators of effect size. Prison programs producing the greatest reductions in misconduct were also associated with larger reductions in recidivism. The magnitudes of various indices of treatment effect size with respect to misconducts were remarkably similar to results in the correctional treatment literature where community recidivism is the criterion. Recommendations are made that will assist prison authorities to manage prisons in a safe and humane manner.

Keywords: prison misconduct reduction; correctional treatment; inmates; prisoners; meta-analysis

The last quarter of the 20th century chronicled an unprecedented surge in prison populations, particularly in the United States, where the incarceration rate has increased in state and federal prisons fourfold in the past 25 years (Lawrence & Travis, 2004). The most recent data from the U.S. Bureau of Justice Statistics report that just more than 2,200,000 offenders were incarcerated at year end 2003 (U.S. Department of Justice, 2004). Scholars have suggested two

AUTHORS' NOTE: *The authors greatly appreciate the ongoing support of Brian Grant for our research endeavors in this area. This project was funded by the Addictions Research Centre, Correctional Service Canada (Contract No. 21120-1-6607). Correspondence concerning this article should be addressed to Sheila French via e-mail at q2gzs@unb.ca.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 33 No. 2, April 2006 185-218
DOI: 10.1177/0093854805284406

primary explanations to account for the growth (e.g., Cullen, Van Voorhis, & Sundt, 1996). First, cohorts of the baby boom were passing through the corrections system. Second, policies stemming from the "get tough" on crime campaign (e.g., mandatory minimum sentencing, parole restrictions, three strikes laws mandating life terms for third offences, War on Drugs) led to a greater number of offenders receiving prison terms and for longer sentences. The increase in incarceration has not been unique to the United States because other countries, albeit for different reasons, have also increasingly relied on incarceration as a method of social control (Cayley, 1998; see also Walmsley, 2003).

The burgeoning inmate populations as well as other factors (e.g., increased programming, inmate transfers, visits) have increased inmate movement and social contact in prisons, which, in turn, has had iatrogenic consequences for prison managers. The consensus among penologists is that the modern prison has become more difficult to manage (DiIulio, 1991; Gendreau, Tellier, & Wormith, 1985; Toch, 1988, Wright, 2000). Partly in response to these developments, a voluminous prison management literature emerged over the past 30 years (e.g., Davies & Burgess, 1988; DiIulio, 1987). Research publications now consist of at least 500 recommendations aimed at making prisons safer, more humane environments (Gendreau & Keyes, 2001).

Within this management literature, prison antisocial behavior, typically operationally defined in terms of misconducts (e.g., physical assaults, theft, disobeying an order; see Gendreau, Goggin, & Law, 1997), has been viewed as the major outcome variable of concern. In the judgement of some researchers in the field, misconduct behavior has reached a critical level in prison; moreover, the situation may well be worsening (Armstrong, 2002; Byrne & Brewster, 1993; Gendreau, Agee, & French, 2004). Although evidence is persuasive that reducing the prevalence of misconduct behavior may result in enormous savings in human and cost-benefit terms (Gendreau & Keyes, 2001; Lovell & Jemelka, 1996), suggestions as to how to achieve this goal have been diverse. In this regard, three general strategies for reducing misconducts have predominated. Each strategy is reviewed here, and the evidence for its respective merit is summarized.

GET TOUGH

"Get tough" advocates of prison reform have suggested a return to "no frills" prisons (Corcoran, 1993; Nossiter, 1994). Such prisons should provide fewer services (e.g., educational, health care, legal, recreational), and they should also encourage the use of solitary confinement as well as lash and chain gangs (Corcoran, 1993; Johnson, Bennett, & Flanagan, 1997; Nossiter, 1994). A modern-day exemplar of a no-frills prison has been the Pelican Bay model in California, where prisoners live under conditions of complete isolation, and on those infrequent occasions when they are allowed outside their cells (e.g., for a medical visit), mechanical restraints are applied (see Haney, 2003). Empirically, there have been virtually no program evaluation studies that have assessed whether no-frills prisons have been effective in reducing misconduct. Results have indicated either little effect (Bidna, 1975) or reductions (Holt & Phillips, 1991; Ralph & Marquart, 1991) in misconducts. Regression to the mean and local history threats to validity may have confounded results from the latter two studies.

PRISON MANAGEMENT/SITUATIONAL CONTROL STRATEGIES

Contributions in this area consist of a highly disparate group of strategies encompassing climatic factors, crowding, food services, inmate personal control over their environments, presumptive parole dates, prison design, prison turnover, racial and gender composition of prisons, staff-prisoner ratios, and visits.

The methodological and theoretical caliber of the investigation in this area is another matter. The great majority of these studies have relied on anecdotal testimonials or weak, quasi-experimental designs based on convenience samples of incarcerates (see Wortley, 2002). Much of the literature has been atheoretical or based on commonsense notions of what may work in reducing misconduct behavior.

Recently, a potentially useful and parsimonious approach to conceptualizing this assemblage has contended that prison-control strategies should be viewed as attempts to minimize situational opportunities for antisocial behavior (i.e., situational crime control; see Felson & Clarke, 1998). Regrettably, Wortley (2002) uncovered a mere handful of studies that explicitly framed their interventions within this model and assessed their value using, at a minimum, quasi-experimental designs.

Given the above limitations, those prison strategies that have provided very tentative evidence as to their utility (but are in dire need of replication) are as follows: (a) climate control (Atlas, 1984; Haertzen, Buxton, Covi, & Richards, 1993), (b) single-cell accommodation (O'Donnell & Edgar, 1996), (c) podular designs wherein correctional officers manage mini prison units that encourage more face-to-face contact and bottom-up decision making (Bayens, Williams, & Smykla, 1997; Senese, 1997; Williams, Rodeheaver, & Huggins, 1999), (d) personal control by inmates over their environment such as lighting and telephone usage (La Vigne, 1994; Wener & Olsen, 1978), (e) low inmate turnover (Porporino & Dudley, 1984); and (f) proper diet (Gesch, Hammond, Hampson, & Eves, 2002; Schoenthaler, 1983a, 1983b). The Gesch et al. (2002) study has arguably the strongest design of any in the situational crime control literature. The present authors calculated the effect size between supplemented diet and misconduct in this study to be approximately r = .10 (i.e., a 10% reduction in misconduct).

Finally, one strategy, which has received by far the most attention, deserves further comment. The effect of crowding, typically defined as spatial density (actual prison population divided by maximum design capacity), at first glance would appear to be a significant contributor to prison unrest. This topic has been the subject of two meta-analyses that have provided an estimate of the magnitude of the crowding effect on misconduct rates. The first, by Bonta and Gendreau (1990), reported a mean Pearson r of .07 for 11 effect sizes between crowding and misconduct. The second study offered a more precise estimate of the crowding effect (r = .03) for 146 effect sizes (Gendreau et al., 1997). After considering sample size, the weighted estimate of Pearson r (Z^+) for the second study was .07, with a 95% confidence interval (CI) of .05 to .09. These authors suggested that it was likely that the rate of inmate turnover, type of inmate management style, and program availability mediated the crowding relationship with misconducts. They proposed that inmate perceptions of feeling overcrowded, not the spatial density of their living space, would be a better predictor of misconducts (see also Walkey & Gilmour, 1984).

TREATMENT PROGRAMS

Of the several hundred management recommendations published for maintaining good order in prisons, the most frequently endorsed single item has been the provision of treatment programs (Gendreau & Keyes, 2001). Surprisingly, despite the emergence of the rehabilitative agenda in corrections in recent years and the accumulation of a large database on "what works" with offender recidivism as the criterion (cf. Andrews & Bonta, 2003; Cullen & Gendreau, 2000), evaluations of the effect of treatment programs on prison misconduct have been rare. As a case in point, Wortley (2002) uncovered only about a dozen studies, some of them based on aggregate data, that had some sort of assessment of program activities (broadly defined) and misconducts. This is an unfortunate situation because the correctional treatment literature, in which program effects on recidivism are the criterion, has much knowledge to offer the prison misconduct literature. Chiefly, this literature has generated three important moderators of treatment effectiveness that have become known as the principles of effective interventions—behavioral programs that target the criminogenic needs of higher risk offenders—that have reported impressive reductions in recidivism (Andrews & Bonta, 2003; Andrews, Dowden, & Gendreau, 1999; Andrews et al., 1990). For example, programs based on the above principles have produced effect sizes of r = .35 and .17 for community and prison-based programs, respectively (see Andrews et al., 1990). Interpreting r using Rosenthal's (1991) binomial effect size display (BESD) statistic, this represents reductions in recidivism of 35% and 17% (see also Cullen & Gendreau, 2000).

Some penologists have asserted that prison misconduct behavior is a reasonable proxy for criminal behavior in the community (Gendreau et al., 1997; Hill, 1985; Homant & Witkowski, 2003; Schnur, 1949; Zamble & Porporino, 1988). That is, the propensity for antisocial behavior on the part of an offender cuts across social situations. Thus, an empirical demonstration that prison treatment programs, founded on some of the principles noted above, could produce roughly equivalent reductions in the case of prison misconducts (e.g., 20% to 30%) would have profound implications for prison managers.

Keyes (1996; see also Gendreau & Keyes, 2001) attempted to address this issue by conducting a quantitative review of the outcome literature relevant to the treatment of prison misconducts. Toward this end, he collected 33 effect sizes, 15 of which were based on behavioral strategies and 18 that were not (e.g., nondirective counseling). The effect size estimate for treatment programs regardless of type was r = .06. The effect size estimate for the 15 behavioral and 18 nonbehavioral treatments were r = .17 and r = –.02, respectively. Seventy-one percent of the time behavioral treatments produced larger reductions in misconducts. Effect sizes were also larger for stronger designs and juvenile samples. Keyes (1996) did not uncover enough information in the studies to code for the targeting of criminogenic needs and offender risk level.

Subsequently, Morgan and Flora (2002) reported on 13 effect sizes for prison group psychotherapy programs. It was apparent from their study that behavioral or cognitive-behavioral approaches tended to produce higher effect sizes, but a precise estimate was not available as the results reported in their comparisons encompassed other outcomes besides misconduct. Our estimate of their effect size, in terms of mean r, was .21 for misconduct outcomes. As with Keyes (1996), Morgan and Flora noted the vast amounts of missing information (e.g., level of therapist training, treatment duration, facility security level, offender assessment) in the studies they collected.

THE PRESENT STUDY

The purpose of the present investigation was twofold. The first and primary objective, and the one central to the interests of prison managers, was to provide a precise estimation, hopefully on a much-expanded database, of the effects of treatment in general and the contribution of the principles of effective treatment—behavioral treatments targeting the criminogenic needs of higher risk offenders—on prison misconducts.

In addition, the influences of two other variables on outcome were examined. The first set refers to the therapeutic integrity of programs, which has been broadly defined as how well a program is maintained as well as the skill level of the therapists and the quality of their therapeutic practices (Andrews & Bonta, 2003; Gendreau & Ross, 1979).

This investigation endeavored to code studies for program therapeutic integrity using an actuarial measure designed for this purpose (Correctional Program Assessment Inventory [CPAI] 2000; Gendreau & Andrews, 2001), which is discussed shortly in more detail. The other variables assessed were suggested by various meta-analyses of offender treatment (cf. Andrews & Bonta, 2003), and they included age, design strength, experimenter involvement, treatment duration, and housing of offenders away from the general population during program participation.

The second goal was to assess whether the results of prison misconduct studies have viable long-term consequences. The latter issue is crucial, for if the proxy "theory" noted previously has substance, then a positive relationship should be discovered between the degree to which a treatment program reduces misconducts and eventual recidivism.

METHOD

LITERATURE SEARCH

Using library abstracting services (PsycINFO, Medline, Academic Search Elite), a literature search was conducted for studies examining prison treatment programs/interventions and their relation to inmate misconduct. Key words used in this search included the names of various psychotherapeutic treatment methods (e.g., behavior modification, cognitive intervention, group therapy) as well as common terms used for offender misconduct (e.g., misconduct, antisocial behavior, behavioral infractions). Also, using the ancestry method, reference lists from each study served as additional sources for relevant literature not produced by the electronic search.

SAMPLE OF STUDIES

To be included in the meta-analysis, studies were required to use a randomized or comparison control group design and to contain sufficient data to calculate an effect size (i.e., Pearson r or phi coefficient) between treatment and criterion. In each study, results from the largest sample, longest follow-up period, and the most serious outcome were recorded. Outcomes included were violent misconduct, nonviolent misconduct, misconduct unspecified, and institutional adjustment rating/scale indices. A study could contribute more than one effect size if the treatment and/or control groups were distinct as to their content/composition (see Andrews et al., 1990). Recidivism data were also recorded when available.

CODING OF STUDIES

The coding guide consisted of 100 coding items, which are available from the first author. The seven general coding categories, with examples of their subcomponents, were as follows:

1. study/author characteristics (e.g., type of publication, author affiliation, publication year),
2. institutional factors (e.g., security level, institution type, location),
3. sample variables (e.g., adult/juvenile, race, gender),
4. research design (e.g., type of design, attrition, sample size, length of follow-up),
5. treatment descriptors (e.g., type of treatment and control groups, treatment dosage),
6. therapeutic integrity items, and
7. effect size descriptors (e.g., type of outcome, calculated effect size).

For item 5, treatment type was coded into one of four categories: behavioral (i.e., radical behavioral, social learning, cognitive behavioral, or punishment), nonbehavioral (e.g., nondirective therapy, psychodynamic, group milieu), educational/vocational, and a nonspecified grouping.

Therapeutic integrity was assessed using the CPAI-2000 (Gendreau & Andrews, 2001). The CPAI-2000 was designed to assess the quality of correctional treatment programs. The measured items were derived from narrative reviews and meta-analyses of the offender treatment literature and from the clinical wisdom of some of the leading practitioners in the field (cf. Gendreau, 1996; Gendreau & Goggin, 1997). The CPAI-2000 has eight dimensions: organizational culture, program implementation/maintenance, management/staff characteristics, client risk/need practices, program characteristics, dimensions of core correctional practice, interagency communications, and evaluation (Gendreau & Andrews, 2001). The measure has been applied extensively to offender treatment programs (Gendreau, Goggin, & Smith, 2001; Matthews, Hubbard, & Latessa, 2001) and has subsequently become a standard tool for "evaluability assessments" in corrections (Van Voorhis & Brown, 1996, p. 6).

Nesovic (2003) has provided a detailed assessment of the psychometric properties of a 1996 version of the CPAI that predated the CPAI-2000. It was found that the measure had high interrater reliability (*kappa* = .84), high face and content validity, and internal consistency estimates ($\alpha = .93$) indicating the measure is a homogeneous scale. Of particular importance is that the measure has strong predictive validities ($r = .46$) for recidivism based on a meta-analysis of the offender treatment literature. Subsequently, Lowenkamp (2004) assessed 38 community treatment programs on the CPAI and correlated the scores with recidivism (program outcomes were based on comparison group designs). Depending on the comparisons being made, the CPAI predicted recidivism in the range of $r = .41$ to .60.

For the present study, 36 items were abstracted from the following four sections of the CPAI-2000: management/staff characteristics, client risk/need practices, program characteristics, and core correctional practices. These items were selected on a rational basis by the second author to serve as an index of therapeutic integrity for treatment studies included in this meta-analysis. Treatment strategies that received a tally between 0 and 4 on this 36-item index were designated as having a *low* level of therapeutic integrity. Strategies scoring between 5 and 9 were assigned to a *medium* level, and a register of 10 or more indicated a *high* level.

All of the studies to be included in the meta-analysis were coded by the first author, who had received prior training in coding from two experienced meta-analysts. Subsequently, the second author (who has had substantial experience with meta-analytic research) independently coded a randomly chosen number (11%) of studies. The results were compared across coders using the Yeaton and Wortman (1993) formula for index of agreement [Σ(agreements) / Σ(agreements + disagreements)]. The agreement index (all variables in the coding guide included) was 88%.

EFFECT SIZE CALCULATION

Phi coefficients (Φ) were calculated using contingency tables for those studies providing information with respect to frequency or proportion of posttreatment misconduct (and recidivism, if applicable) for both experimental and control groups. When statistics other than *r* were reported (i.e., *F*, *t*, χ^2, *p*), the appropriate formula for conversion to Pearson *r* was used (Rosenthal, 1991). In cases in which an author report of nonsignificance or a *p* value greater than .05 was the only significance information provided, a Pearson *r* of .00 was assigned (Rosenthal & DiMatteo, 2001).

EFFECT SIZE MAGNITUDE

The metrics used as estimates of the magnitude of misconduct/institutional adjustment effect sizes for each category of interest (e.g., behavioral treatment programs) were mean Pearson *r* and weighted estimates of Pearson *r* (Z^+), along with associated 95% confidence intervals (CI). Although the CI can be interpreted as a significance test based on whether the interval includes zero, significance testing was not emphasized in this study (see Gendreau, 2002; Schmidt, 1996). Rather, the CI was used to emphasize (a) the degree to which there was agreement among study variables (i.e., if CIs overlapped they were assumed to be estimating the same population parameter), and (b) the uncertainty of effect size estimates, which was judged by noting the width of the CI (see Gendreau, Goggin, & Smith, 2000; Schmidt, 1996).

To calculate the combined weighted effect size estimate (Z^+) for each domain (e.g., behavioral treatment programs), each recorded *r* value within that category was first transformed to Z_r using Fisher's table of Z_r values (Rosenthal, 1991). This was necessary to allow for combination and comparison of *r* values using a less skewed, more normalized distribution (Rosenthal, 1991). For example, a study reporting a Pearson *r* of .32 between participation in a behaviorally based program and institutional misconduct would have been transformed into a Z_r value of .33. Subsequently, each Z_r was then weighted by the size of its associated sample minus three (Hedges & Olkin, 1985). Thus, for the behavioral program example, if the sample size were 200, the weighted Z_r value ([Z_r][n – 3]) would have been 65.01. With this information, the combined effect size estimate (Z^+) per category was then calculated by dividing the sum of the weighted Z_r values by the sum of the sample size minus three (Σ[n – 3]).

ASSESSMENT OF SAMPLING ERROR

Using the procedure recommended by Hunter and Schmidt (1990), the degree to which sampling error contributed to variability in this collected distribution of effect sizes was assessed by partitioning variance into two components: observed variance and variance because of sampling error. According to Hunter and Schmidt's "75% rule" (p. 414), if the variance accounted for by sampling error is less than 75% of the total variance observed, exploration of moderating variables is warranted. However, if researchers specify a priori hypotheses, a search for relevant moderators may be carried out even if the 75% rule has been exceeded.

To calculate observed variance in this distribution of effect sizes, the following method was used. First, the standard deviation for each effect size was calculated using the following equation:

$$SD_r = \sqrt{(1 - r^2)^2 / N - 1} \quad (1)$$

where r was each individual effect size and N was the associated sample size per effect (Hunter & Schmidt, 1990).

The results were summed and averaged across all effect sizes to yield a standard deviation for the distribution. This average was then squared to obtain the observed variance of the effect sizes (S_r^2).

To determine the proportion of observed variance attributed to sampling error (S_e^2 / S_r^2), error variance was first calculated using

$$S_e^2 = \sum_{i=1}^{N} n_i [(1 - \bar{r}^2)^2 / n - 1] / \sum_{i=1}^{N} n_i \quad (2)$$

where N was the number of effect sizes, n_i was the sample size associated with each effect size, and $\bar{r}^2$ was the square of the mean of the effect sizes.

ELIMINATING OUTLIERS

In meta-analysis, the distribution of effect sizes within a domain is assessed to determine homogeneity (i.e., the dispersion of effect sizes around their mean is no greater than that expected from sampling error alone).

Variability was examined using the *Q* statistic (Hedges & Olkin, 1985). Calculation of the *Q* statistic proceeded as follows. For each effect size within each category (e.g., treatment type), a *q* value was calculated using the formula:

$$q = [(n - 3)(Z_r - MZ_r)] \quad (3)$$

where n was the sample size per effect size, Z_r was the standardized r per effect size, and MZ_r was the weighted mean Z_r for each treatment category. These q values were then summed for each type of treatment, yielding Q. To evaluate its significance, the Q for each treatment category was assessed using the critical value of a χ^2 distribution with $(k - 1)$ degrees of freedom, where k is the number of effect sizes in the category. If significant heterogeneity was indicated, outlying weighted effect sizes ($[Z_r][n - 3]$) positioned above or below the mean of each treatment category by two or more standard deviations were identified and removed. This process was repeated until nonsignificant heterogeneity was achieved or until the originally obtained Q was reduced by 50% (Bonta, Law, & Hanson, 1998).

COMMON LANGUAGE EFFECT SIZE INDICATOR

McGraw and Wong's (1992) common language effect size indicator (CL) was used to determine the practical utility of the categories of interest in each comparison. The CL statistic converts an effect size into a probability that a score (i.e., an effect size) sampled from the distribution of one group will be larger than a score sampled from that of another. Note the CL statistic cannot be calculated for a Z^+ estimate of effect size, as this statistic does not produce a standard deviation.

FAIL-SAFE ESTIMATION

A fail-safe estimate was used to provide an index of how many previously uncovered or "file drawer" effect sizes would be required to alter the obtained results. For example, an index of the number of effect sizes (r = .00) needed for a treatment type of greater efficacy (Treatment A) in the reduction of misconduct to approach that equal to one of lesser efficacy (Treatment B) was calculated using the following formula:

$$N_{fs} = [(K_A(r_A - r_B)) / (r_B - r_{A=0})] \quad (4)$$

where $r_{A=0}$ indicates a null effect for the more efficacious treatment type (see Gendreau, Goggin, & Smith, 2002).

As applied to this meta-analysis, assume that the mean effect size for Treatment Type A was $r_A = .35$ ($k_A = 50$) and that of Treatment Type B was $r_B = .25$ ($k_B = 40$). An estimate of the number of Type A treatments with r = .00 required to negate its supremacy using the above formula is 20. That is, 20 additional Treatment Type A effect sizes, each with a magnitude of r = .00, would have to be located to conclude that the two treatments were at parity.

RESULTS

EFFECT SIZE/STUDY CHARACTERISTICS

Sixty-eight studies were collected, from which 104 effect sizes (r) between various types of treatment programs and prison misconduct were calculated. The majority of articles provided very little information as to their study characteristics, which are reported here in terms of effect size percentages. We itemize those coding items for which information was available for at least 50% of effect sizes. Also, unless otherwise noted, subcomponents of coding items are reported only when the frequency was 20% or greater.

Eighty-one percent of effect sizes were reported in journal articles. The studies collected spanned the years 1952 to 2003. The percentages of effect sizes produced by decades were: 1952-1960 (8%), 1961-1970 (18%), 1971-1980 (30%), 1981-1990 (25%), and 1991-2003 (19%). Fifty percent of effect sizes were associated with academically affiliated authors, and 19% from government affiliations. Eighty-two percent of effect sizes originated from U.S. prisons.

Seventy-three percent of effect sizes came from male samples, and 8% came from female samples. The remaining 19% came from studies with mixed samples or in which gender was not specified. Forty percent and 49% were from adult and juvenile samples, respectively. Thirty-nine percent of effect sizes were recorded from studies with strong research designs. Sample sizes ranged from 5 to 1,478 incarcerates for treatment and comparison groups. The percentages of effect sizes by length of follow-up were: less than 6 months (45%), 6 to 11 months (16%), 12 months or longer (17%), and follow-up information not reported (22%).

Seventy-five percent of effect sizes could be attributed to a specified treatment modality. Thirty-eight percent of these were designated behavioral in nature (e.g., 9.5% cognitive behavioral, 4% radical behavioral, 7% social learning, 14% mixed), and 30% were nonbehavioral (e.g., 7% group milieu, 5% nondirective, 5% diet). Forty-eight percent of effect sizes were calculated from programs of at least 3 months duration, with 20% stemming from programs that required more than 50 treatment hours. The majority of effect sizes (75%) were derived from studies using "no treatment" comparison groups, whereas 8% and 6% came from unspecified alternate treatment and wait-list comparison groups, respectively. Programs for which it was determined that no criminogenic needs were targeted accounted for 17% of effect sizes. Thirty-nine percent of effect sizes were from programs that had addressed one to two criminogenic needs, whereas 23% came from programs in which it was ascertained that treatment targets consisted of three to eight criminogenic needs. Programs rating *low* (0-4), *medium* (5-9), and *high* (10+) on the 36-item scale of program therapeutic integrity accounted for 36%, 23%, and 10% of effect sizes, respectively.

The frequency of missing information in terms of percentages of effect size ranged from 51% to 100% for the following categories: (a) institutional characteristics (e.g., institution type, security level, location of inmates during programming, population size, crowding index, treatment orientation, prison climate), (b) offender characteristics (e.g., race, risk level, risk assessment, misconduct history), (c) program director and staff characteristics (e.g., qualifications, experience, education, training, level of therapeutic skills), (d) treatment characteristics (e.g., specific responsivity, risk level, program manual,

TABLE 1: Mean Effect Sizes for Misconduct by Treatment Type

Treatment Type	k	N	Mr *(*SD*)*	CI_r	Z^+	CI_{z+}	Q
Behavioral	40	5,809	.26 (.24)	.18 to .34	.39	.36 to .42	505.77*
With outliers eliminated	39	2,853	.25 (.24)	.17 to .33	.17	.13 to .21	233.69[b]*
Nonbehavioral	31	7,629	.10 (.21)	.02 to .18	.05	.03 to .07	128.64*
With outliers eliminated	28	5,150	.11 (.21)	.03 to .19	.05	.02 to .08	94.45[a]*
Educational/ vocational	8	1,117	.02 (.19)	–.14 to .18	.00	–.06 to .06	21.07*[a]
Unspecified	25	6,912	.03 (.14)	–.03 to .09	.07	.05 to .09	138.41*
With outliers eliminated	23	3,639	.02 (.13)	–.04 to .08	.02	–.01 to .05	45.67[b]*

Note. k = effect sizes per treatment type; N = offenders per treatment type; Mr(SD) = mean Pearson r and standard deviation between treatment type and misconduct; CI_r = 95% confidence interval about mean Pearson r; Z^+ = weighted r value that accounts for the influence of sample size; CI_{z+} = 95% confidence interval about Z^+; Q = statistic evaluated to determine heterogeneity of effect sizes within category.
a. Although *Q* was significant, no outliers were identified.
b. *Q* is less than one half original *Q* value; therefore, no further outliers were removed.
*$p < .05$.

therapeutic practices), (e) the criterion (i.e., type of misconducts and recidivism), and (f) the precise metric used (e.g., *r*, *t*, *F* values).

OVERALL TREATMENT EFFECTS ON MISCONDUCT

One hundred and four effect sizes, representing 21,467 inmates, were extracted from 68 studies. The mean effect size was $r = .14$ (CI = .09 to .18), and after weighting by sample size, the weighted mean effect size (Z^+) was .14 (CI = .13 to .15). Applying the Hunter and Schmidt (1990) "75% rule" to the database, 34% of observed variance in the distribution of effect sizes was attributed to sampling error. This indicated that a search for moderating factors (e.g., treatment type, criminogenic needs targeted) was warranted. Offender risk level was not assessed as so few effect sizes ($k = 7$) contained relevant information.

TREATMENT TYPE: EFFECT ON MISCONDUCT

The mean effect sizes for misconduct categorized by treatment type are summarized in Table 1. For example, according to row 1 in Table

1, there were 40 behavioral treatment effect sizes reported involving 5,809 offenders. The mean *r* was .26, with a CI of .18 to .34. Thus, there was a 26% reduction in misconduct according to the BESD statistic (Rosenthal, 1991). After weighting by sample size and number of effect sizes, the weighted effect size (Z^+) was .39, with a CI of .36 to .42.

By comparison, the mean *r* for the 31 nonbehavioral treatment effect sizes (N = 7,629) was .10, with an associated Z^+ of .05. The 95% CIs for mean *r* and Z^+ for the behavioral treatment group did not overlap with the CIs for the nonbehavioral, educational/vocational, and unspecified treatment categories. According to the CL statistic, behavioral treatments produced greater reductions in misconduct 69% of the time versus nonbehavioral treatments, 77% of the time versus educational/vocational strategies, and 80% compared to unspecified treatments.

Fail-safe analysis indicated that in order for the behavioral treatment mean effect size to be reduced to the effect size level for nonbehavioral treatment, 64 additional behavioral treatment studies, all with an $r = .00$, would need to be located.

Testing for effect size heterogeneity using the Hedges and Olkin (1985) procedure warranted elimination of 1, 3, and 2 effect sizes for the behavioral, nonbehavioral, and unspecified categories, respectively.

Table 1 also depicts the effect size estimators for *r* and Z^+ with outliers eliminated. After removal of the one outlier ($r = .53$, $N = 2,956$), the mean *r* for behavioral treatments became .25 (CI = .17 to .33). After weighting, there was a noticeable reduction in Z^+ (reduced from .37 to .17, CI = .13 to .21).

Nonbehavioral treatments ($N = 5,150$, $k = 28$) resulted in a mean *r* of .11 (CI = .03 to .19), with a weighted mean effect estimate (Z^+) equal to .05 (CI = .02 to .08). There was a minimal degree of overlap in the CIs between behavioral and nonbehavioral treatment groups about mean *r* but not for Z^+.

Turning to comparisons with outliers removed, the CL index indicated larger mean *r* values for behavioral treatments 67% of the time versus nonbehavioral treatments, 76% of the time versus educational/

TABLE 2: Mean Effect Sizes for Misconduct by Number of Criminogenic Needs Targeted

Criminogenic Needs	k	N	Mr *(SD)*	CI_r	Z^+	CI_{z+}	Q
Three to eight	24	4,586	.29 (.23)	.19 to .39	.47	.44 to .50	322.14*
With outliers eliminated	23	1,630	.28 (.23)	.18 to .38	.25	.20 to .30	120.11[c]*
One to two	41	5,063	.16 (.23)	.08 to .23	.09	.06 to .12	199.07*
With outliers eliminated	38	3,750	.16 (.23)	.08 to .24	.06	.03 to .09	152.98[b]*
Zero[a]	18	4,526	.06 (.17)	–.02 to .14	.04	.01 to .07	76.56[b]*

Note. k = effect sizes per category; N = offenders per category; Mr (SD) = mean Pearson r and standard deviation between category and misconduct; CI_r = 95% confidence interval about mean Pearson r; Z^+ = weighted r value that accounts for the influence of sample size; CI_{z+} = 95% confidence interval about Z^+; Q = statistic evaluated to determine heterogeneity of effect sizes within category.
a. Studies targeting no criminogenic needs targeted from zero to six noncriminogenic needs.
b. Although *Q* was significant, no outliers were identified.
c. *Q* less than one half original *Q* value; therefore, no further outliers were removed.
$*p < .05$.

vocational strategies, and 79% of the time versus unspecified treatment types. Results of the fail-safe analysis indicated that an additional 50 behavioral treatment effect sizes with $r = .00$ would be needed for the behavioral treatment effect to equal that of the nonbehavioral treatment.

CRIMINOGENIC NEEDS TARGETED

There were a total of 83 effect sizes involving 14,175 inmates, for which the number of criminogenic needs targeted could be determined. The results of this analysis are detailed in Table 2.

To facilitate comparison, categories were created based on the number of needs targeted. For example, reading across the first row, it can be seen that there were 24 effect sizes in which three to eight criminogenic needs were targeted ($N = 4{,}586$), the mean r for this category was .29, and weighting by sample size substantially increased the mean effect size ($Z^+ = .47$). By comparison, those studies targeting one to two or no criminogenic needs produced lower effect sizes. The

CIs for these latter two categories overlapped. Even though the weighted mean effect size for the three to eight classification was reduced with outliers eliminated (Z^{+} = .25), it was still much larger than the other two groupings.

According to the CL index, reductions in misconduct for those studies that targeted three to eight criminogenic needs were greater 66% of the time when compared to those that targeted only one to two criminogenic needs, and they had greater effects 79% of the time versus those studies that targeted no criminogenic needs. With outliers removed, the CL percentages were: three to eight versus one to two, 64%; three to eight versus zero, 80%; one to two versus zero, 54%.

Fail-safe analysis suggested that another 20 treatment effect sizes of r = .00 targeting three to eight criminogenic needs would be required for the mean r of that level to approach the mean effect of treatments targeting one to two needs, and a further 92 effect sizes of r = .00 would be required for it to be at parity with the mean effect of programs targeting no needs.

THERAPEUTIC INTEGRITY

Averaging across all effect sizes for which a therapeutic integrity score could be determined (k = 72), the mean Pearson r with misconduct was .19 (CI = .13 to .25), with a weighted average effect (Z^{+}) of .26 (CI = .24 to .28).

The mean r and Z^{+} effect sizes for the three levels of therapeutic integrity are reported in Table 3. The mean effect with misconduct for treatments of high therapeutic integrity (r = .38) was superior to the mean effect for treatments with medium (r = .20) and low (r = .13) levels. The ordering of results remained the same after weighting for sample size as well as when outliers were removed. The CIs associated with mean r for the three levels overlapped. After weighting and also after the removal of outliers, the CI for the high level did not overlap with the other two levels.

CL calculation favored effect sizes in the high category when compared to effect sizes in both the medium and low level 69% and 75% of the time and 70% and 73% with outliers removed.

TABLE 3: Mean Effect Sizes for Misconduct by Level of Therapeutic Integrity[a]

Therapeutic Integrity	k	N	Mr *(*SD*)*	CI_r	Z^+	CI_{z+}	Q
High	10	420	.38 (.30)	.16 to .60	.47	.37 to .57	63.50[b]*
Medium	24	4,841	.20 (.20)	.12 to .29	.42	.39 to .45	480.49*
With outliers eliminated	23	1,885	.19 (.19)	.11 to .27	.14	.09 to .19	54.08[c]*
Low	38	5,060	.13 (.23)	.05 to .20	.08	.05 to .11	202.74*
With outliers eliminated	32	2,708	.16 (.21)	.08 to .23	.08	.04 to .12	84.53[c]*

Note. k = effect sizes per category; N = offenders per category; Mr (SD) = mean Pearson r and standard deviation between level of therapeutic integrity and misconduct; CI_r = 95% confidence interval about mean Pearson r; Z^+ = weighted r value that accounts for the influence of sample size; CI_{z+} = 95% confidence interval about Z^+; Q = statistic evaluated to determine heterogeneity of effect sizes within category.
a. Levels of therapeutic integrity were calculated from the Therapeutic Integrity Descriptors section of the coding guide. Effect sizes in the *low* category had positive responses on a total of 0 to 4 of the 33 items, those in the *medium* category earned positive responses for 5 to 9 of the items, and the *high* category included effect sizes earning positive responses on 10 or more items.
b. Although *Q* was significant, no outliers were identified.
c. *Q* less than one half original *Q* value; therefore, no further outliers were removed.
$^*p < .05$.

Fail-safe analysis revealed that a further 19 null effect sizes with high levels of therapeutic integrity would be needed for the effect size of the high-level group to be reduced to that of the low-level group.

MODERATORS OF MISCONDUCT EFFECT SIZE

To explore the effects of other moderators on the treatment type/misconduct relationship, a minimum number of 10 effect sizes were required within each treatment category for further analysis. This condition was met for only the moderator variables of design strength and age.

With regard to design strength, strong experimental designs (k = 41) were those designs employing randomization of participants to treatment and control groups (95% of effect sizes) or when a treatment group was matched with a comparison group on five or more criminal risk factors (5% of effect sizes). Criteria for classification as a weak experimental design (k = 63) were as follows: more than 20% attrition in experimental or control group (22% of effect sizes), comparison group matched on fewer than five criminal risk factors (50% of effect sizes), and postdictive correlation designs (28% of effect sizes).

The mean effect sizes for treatment type after partitioning effect sizes by design strength indicated no consistent support for design strength functioning as a moderator. The respective mean *r* values and CIs for strong and weak design effect sizes from behavioral programs were $r = .36$ (CI = .20 to .52) and $r = .21$ (CI = .12 to .30). On weighting, the Z^+ values were .40 and .39, respectively. Results for nonbehavioral and unspecified treatment programs across both types of designs slightly favored weak designs, but in all cases (mean *r* and Z^+) there was considerable overlap in the CIs.

Similar results were reported for age. Adult samples were those in which offenders were 18 years of age or older, and samples with offenders younger than 18 years of age were designated as juvenile.

There were 10 effect sizes in the behavioral category involving 848 adult inmates. The mean effect for adults was $r = .20$ (CI = .07 to .32), with a weighted mean effect of $Z^+ = .20$ (CI = .13 to .27). In comparison, the mean effect for the behavioral category involving 1,651 juvenile inmates was $r = .27$ (CI = .15 to .38), which was reduced to a weighted mean effect of $Z^+ = .18$ (CI = .13 to .23). There was substantial overlap of the CIs (both *r* and Z^+) for this comparison as well as for the adult-juvenile comparisons of nonbehavioral and unspecified treatment categories.

OTHER RELATIONSHIPS WITH MISCONDUCT EFFECT SIZE

The following variables were not examined as moderators of effect size by treatment type because there were fewer than 10 effect sizes in one or more of the treatment categories: experimenter involvement in the treatment program, separate housing of the treatment sample from the general prison population, length of follow-up, and duration of treatment (i.e., dosage). Their relationship with effect size was examined irrespective of the treatment category. Of a possible 44 comparisons, the CIs did not overlap in the following instances:

1. The Z^+ CIs for offender isolation from the general population while in the treatment program (Z^+ = .33, CI = .31 to .35) versus offenders not housed separately (Z^+ = .05, CI = .02 to .08).
2. The Z^+ CIs for 1- to 2-year follow-up periods (Z^+ = .27, CI = .25 to .29) versus 6- to 11-month periods (Z^+ = .09, CI = .05 to .13), 3- to 5-month periods (Z^+ = .09, CI = .05 to .13), and periods less than 3 months (Z^+ = .00, CI = –.04 to .05).
3. The Z^+ CIs for treatment durations of greater than 6 months (Z^+ = .28, CI = .26 to .30) versus 3- to 5-month durations (Z^+ = .15, CI = .10 to .20) and durations of fewer than 3 months (Z^+ = .15, CI = .09 to .21).

RECIDIVISM OUTCOMES

Twenty-three effect sizes were associated with recidivism outcomes. They were grouped as follows: effect sizes with high reductions in misconduct (k = 12) that reported recidivism outcomes and effect sizes with low reductions in misconduct (k = 11) that reported recidivism outcomes.

Of the 12 high-misconduct-reduction effect sizes, 92% were attributed to programs that were behavioral in nature. Fifty percent of the 12 effect sizes were derived from programs that targeted three or more criminogenic needs, and 92% had therapeutic integrity level scores in the medium to high categories.

In the case of low-misconduct-reduction effect sizes (k = 11), 36% came from programs designated as behavioral, 10% from programs that targeted three or more criminogenic needs, and 36% from programs considered medium to high in therapeutic integrity.

Those effect sizes in the high-misconduct-reduction category (k = 12) were associated with the following mean reduction in recidivism for r (r = .13; CI = –.04 to .29) and Z^+ (Z^+ = .07; CI = .01 to .13). For those effect sizes in the low-misconduct-reduction category (k = 11), there was either an increase in recidivism (r = –.05; CI = –.16 to .07) or a slight decrease (Z^+ = .03; CI = –.02 to .08).

The CL index favored recidivism effect sizes associated with the high category of misconduct effect sizes 71% of the time versus recidivism effect sizes associated with the low category of misconduct effect sizes.

DISCUSSION

Previous meta-analyses (Keyes, 1996; Morgan & Flora, 2002) of the effects for prison-based treatment programs on misconducts amassed a combined total of 32 studies with 46 effect sizes. The present study increased this database to 68 studies that produced 104 effect sizes on a sample of just more than 21,000 inmates. This represents a considerable advance in knowledge cumulation for the purposes of assessing the use of treatment programs for managing prisons in a safe and humane manner.

Before discussing the results, it must be noted that there were serious limitations regarding the quality of information provided by studies in this meta-analysis. This, unfortunately, is a finding consistent with previous meta-analyses on prison research topics (Gendreau et al., 1997; Gendreau, Goggin, & Cullen, 1999). The lack of information for key variables is of major concern because it was impossible to examine potentially important moderator influences on effect-size estimations. In a few cases (e.g., Liau, 1999; Wormith, 1984), authors were contacted for clarification or provision of data. In most cases, however, authors could not be located (i.e., some articles were published 20 to 50 years ago, making location of the original authors virtually impossible). Almost invariably, information was not reported on the prison context (e.g., crowding, institutional climate) within which the treatment programs operated. Essential inmate characteristics such as offender risk level and misconduct history were almost always missing (more than 90% of studies), as was information regarding the quality or therapeutic integrity of the programs themselves. Only 2% of effect sizes came from treatment programs that scored better than 50% in therapeutic integrity (i.e., scores greater than 18 out of 36) using the truncated version of the CPAI-2000. This extremely low estimate was likely because of reporting practices in the journal articles and the fact that the majority of studies were published before much was known about the principles of effective correctional treatment. Only on-site evaluations of program quality will adequately address the importance of this factor. Also noteworthy was the fact that, in about one quarter of the effect sizes, it was impossible to identify the exact nature of the treatment, and the precise outcome metric was not available 52% of the time.

The reporting of "nonsignificance" was frustrating as this practice likely served to underestimate treatment effects. In some cases, there were seemingly modest to strong effects to report, but the authors dismissed their findings because sample sizes were too small to produce the $p < .05$ arbiter of success. One recourse in these cases is to assign an *r* value equal to zero (Rosenthal & DiMatteo, 2001). This may seem harsh, but there are instances in the offender treatment literature in which programs have produced slight increases in recidivism (Andrews et al., 1990), thus a zero estimate is prudent. Finally, one must be mindful of the fact that although it has been shown that prison-misconduct records have reasonable accuracy (Gendreau et al., 1997; Van Voorhis, 1994), there have been situations in which the recording of misconducts by prison authorities has been unreliable (Light, 1990).

Granted these important caveats, the results that emerge from the present study are consistent with results from two previous meta-analyses. Keyes (1996) and Morgan and Flora (2002) reported that behavioral treatments produced reductions in misconducts of approximately $r = .20$. This meta-analysis confirms their results and indicates that the effects of such treatments are more robust ($r = .26$), particularly in the case when effect sizes are weighted by sample size ($Z^+ = .39$). The marked increase in the latter (Z^+) is because of one study by Prendergast, Farabee, and Cartier (2001) with a huge effect size of $r = .53$ on a large sample of 2,956 offenders. This effect size was eliminated as an outlier (see Table 1), which produced very little change in the mean *r* value but caused an obvious reduction of the weighted mean effect size (from $Z^+ = .39$ to .17). The Prendergast et al. study was not classified as a strong design (recall that strong designs were not associated with lower effect sizes) but employed a program that scored in the medium range for therapeutic integrity on the CPAI-2000 and targeted an impressive number of criminogenic needs (e.g., antisocial attitudes, behaviors, associates) that are among the strongest predictors of criminal behavior (Gendreau, Little, & Goggin, 1996).

Some meta-analysts have expressed skepticism about eliminating studies purely on statistical significance testing grounds (Hunter & Schmidt, 1990; Rosenthal & DiMatteo, 2001). By doing so, one runs the risk of eliminating studies of considerable merit be it in terms of sample size, design quality, or quality of the intervention employed (Gendreau, Goggin, & Smith, 2003). In fact, in this meta-analysis, some of the largest weighted effect sizes come from studies by research groups that have had a distinguished history of effective service delivery with offenders (e.g., Goldstein, Glick, Reiner, Zimmerman, & Coultry, 1987; Leeman, Gibbs, & Fuller, 1993) or others that have designed programs with high therapeutic integrity (Guerra & Slaby, 1990; Ollendick & Hersen, 1979). Therefore, for those researchers whose sole concern is not significance testing but the estimation of the magnitude of an effect, the circumspect strategy would be to consider including all effect sizes unless there are compelling reasons not to do so (Gendreau et al., 2003).

The current results are persuasive when it comes to the choice of program for reducing misconducts. The effect size for behavioral programs was $r = .26$, with a 95% CI of .18 to .34. The corresponding results for nonbehavioral and educational/vocational programs, respectively, were $r = .10$ (CI = .02 to .18) and $r = .02$ (CI = –.14 to .18). Sixty-nine and 77% of the time behavioral programs produced a better result than nonbehavioral and educational/vocational programs. Might there be a number of forthcoming or hidden studies that would challenge these findings? The present investigators did not have the resources to solicit studies “hidden” in prison/government files (the most likely place), where the number of which might be significant. Our guess is that it is doubtful that, in the specific case of comparing treatment modalities, the superiority of behavioral programs will be challenged. This study found that the next generation of behavioral programs would have to perform abysmally (64 consecutive effect sizes reported with $r = .00$) to be reduced in effectiveness to that of the nonbehavioral category.

The practical implications of these results for prison managers cannot be stressed too highly. Applying Rosenthal’s (1991) BESD statistic to the behavioral *r* value estimate of effect reported in this study indicates that misconducts can be reduced by about 26% for behavioral program participants or, if referring to the CI, the probability is .95 that the interval (.18 to .34) contains the true population value. This is a result that any prison administrator would eagerly embrace as a useful finding. For those readers with experience working in prisons, even a small reduction (e.g., 10%) in misconducts for a group of disruptive inmates might mean the difference between having a chaotic prison environment and having one that is coping adequately with the usual pressures. Furthermore, six-figure cost savings can result from a modest reduction in misconducts for some prisons (cf., Lovell & Jemelka, 1996).

Admittedly, the results of comparisons of effect sizes across different studies should be interpreted with caution, but an especially interesting finding is worth comment here. The effect sizes produced by this study are congruent with those in the correctional treatment and recidivism literature. Consider the following parallels between the two literatures. The mean reduction in misconducts for behavioral programs in this study was $r = .26$, and in a meta-analysis for similar programs and recidivism outcomes in the community, the result was $r = .23$ (Andrews & Bonta, 2003). The misconduct mean effect size for those studies targeting one or more criminogenic needs in this meta-analysis was $r = .21$ (from Table 2, averaging categories 3-8 and 1-2), which is identical to the corresponding literature on recidivism (Andrews & Bonta, 2003). The measure of therapeutic integrity used in this study produced a mean reduction in misconduct of $r = .29$ (from Table 3, combining the high and medium categories). The mean r value for the indices of therapeutic integrity reported by Andrews and Bonta (2003) for recidivism was $r = .23$. The moderator analysis also produced results consistent with reports in the offender treatment/ recidivism literature. That is, greater treatment dosage and the offender being housed apart from the general population while in treatment were associated with higher effect sizes (cf., Cullen & Gendreau, 2000; Gendreau, 1996), whereas research design was not (Andrews et al., 1990).

There was, however, some incongruence between the present findings and those in the correctional treatment and recidivism literature. In this study, experimenter involvement with the program was not found to be associated with higher effect sizes, as was reported by Andrews and Bonta (2003). This failure to replicate may have been because of sampling error or the lack of information in study methods sections. Unlike Keyes's (1996) findings, age and research design were not correlated with effect size. This may have been because of sampling error (Keyes reported on only 33 effect sizes) or to the fact that the CIs for the therapeutic integrity effect sizes for each of the younger versus older samples and the stronger versus weaker research designs, overlapped in this meta-analysis. This finding raises the intriguing notion that when moderator differences are not found for age and research design in future meta-analyses, one possible underlying factor might be that the therapeutic integrity of the programs being studied supersedes other factors.

The finding that programs that were most effective in reducing prison misconducts also generated lower recidivism rates (r = .13, CI = –.04 to .29) in the community was an especially gratifying result. It reinforces the view that prison misconduct behavior is a reasonable proxy for antisocial behavior in the community (e.g., Gendreau et al., 1997; Hill, 1985; Zamble & Porporino, 1988). This fact ably demonstrates the beneficial long-term consequences of providing quality treatment programs, a management policy most frequently endorsed by “experts” in corrections (see Gendreau & Keyes, 2001). This effect size (r = .13) is also typical of the effects of prison behavioral programs for recidivism, which are about half of the magnitude of similar community programs (e.g., Andrews et al., 1990), presumably because treatment gains dissipate while an offender remains incarcerated. Caution is required in regards to the results in this aspect of our study because the width of the 95% CI is considerable, and the interval includes zero as well as negative effect size estimates. Thus, replications of the positive point estimate result (r = .13) of programs that reduce misconducts and their effect on recidivism are necessary.

In summary, this meta-analysis, based on a large sample of incarcerates, indicated that, on average, prison-based behavioral programs produce large reductions in misconducts that may carry over into reductions in recidivism in the community. This result represents a significant advance in knowledge concerning prison management. The results, however, do not mean the case is closed. In fact, we view this meta-analysis as just a first step in the knowledge cumulation process regarding what works in prisons.

First, huge gaps in the literature were revealed. This is an area in which, in our view, research emphasis in the future should be directed less toward meta-analysis and more toward conducting primary studies that carefully define prison context and the criterion used (i.e., type of misconduct). Future research should also employ longer follow-up periods and supply detailed information on offender characteristics such as risk level. With more attention paid to these moderating factors, it is possible that some potent interactions with effect size might be discovered. We are also of the opinion, and we emphasize this point, that the largest treatment gains will only be revealed from onsite evaluations of programs when all of the information is available for scoring on measures of therapeutic integrity. Some prison systems are slowly moving in this direction by accrediting prison treatment programs on the basis of the quality of their interventions (cf., Gendreau, Goggin, French, & Smith, in press). We surmise that the end result of such initiatives, assuming that they are published, will demonstrate how useful well-designed and well-conducted treatment programs are for the purpose of reducing inmate antisocial behavior. Finally, prison policy makers should consider further developing and evaluating those programs based on opportunity-reduction strategies (see Wortley, 2002). The combination of strategies for controlling behavior in the immediate situation with the longer-term benefits of behavioral treatment programs augurs well for establishing sound prison management policies.

REFERENCES

*References marked with an asterisk indicate studies that were included in the meta-analysis.

*Abrams, A. I., & Siegel, L. M. (1978). The transcendental meditation program and rehabilitation at Folsom State Prison. *Criminal Justice and Behavior*, *5*, 3-30.

Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct* (3rd ed.). Cincinnati, OH: Anderson.

Andrews, D. A., Dowden, C., & Gendreau, P. (1999). *Clinically relevant and psychologically informed approaches to reduced re-offending: A meta-analytic study of human service, risk, need, responsivity and other concerns in justice contexts*. Ottawa, Canada: Carleton University.

*Andrews, D. A., & Young, J. G. (1974). Short-term structured group counselling and prison adjustment. *Canadian Journal of Criminology and Corrections*, *16*, 5-13.

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A psychologically informed meta-analysis. *Criminology*, *28*, 369-404.

*Armstrong, T. A. (2002). The effect of environment on the behavior of youthful offenders: A randomized experiment. *Journal of Criminal Justice*, *30*, 19-28.

Atlas, R. (1984). Violence in prison: Environmental influences. *Environment and Behavior*, *16*, 275-306.

*Barratt, E. S., Stanford, M. S., Felthous, A. R., & Kent, T. A. (1997). The effects of phenytoin on impulsive and premeditated aggression: A controlled study. *Journal of Clinical Psychopharmacology*, *17*, 341-349.

Bayens, G., Williams, J., & Smykla, J. (1997). Jail type and inmate behaviour: A longitudinal analysis. *Federal Probation*, *61*, 54-62.

Bidna, H. (1975). Effects of increased security on prison violence. *Journal of Criminal Justice*, *3*, 33-46.

Bonta, J., & Gendreau, P. (1990). Reexamining the cruel and unusual punishment of prison life. *Law and Human Behavior*, *14*, 347-372.

Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin*, *123*, 123-142.

*Brill, R. (1978). Implications of the conceptual level matching model for treatment of delinquents. *Journal of Research in Crime and Delinquency*, *15*, 229-246.

Byrne, J., & Brewster, M. (1993). Choosing the future of American corrections: Punishment or reform? *Federal Probation*, *57*, 3-9.

*Cavior, H. E., & Schmidt, A. (1978). A test of the effectiveness of a differential treatment strategy at the Robert F. Kennedy Center. *Criminal Justice and Behavior*, *5*, 131-139.

Cayley, D. (1998). *The expanding prison*. Toronto, Canada: House of Anansi.

Corcoran, L. (1993). Comfortable prisons push recidivism rates. *Peacekeeper*, *2*, 7-8.

*Craft, M., Stephenson, G., & Granger, C. (1964). A controlled trial of authoritarian and self-governing regimes with adolescent psychopaths. *Journal of Orthopsychiatry*, *34*, 543-554.

*Cullen, E. (1987). Group-based treatments for serious institutional offending. In B. J. McGurk, D. M. Thornton, & M. Williams (Eds.), *Applying psychology to imprisonment: Theory and practice* (pp. 315-327). London: Her Majesty's Stationery Office Books.

Cullen, F. T., & Gendreau, P. (2000). Assessing correctional rehabilitation: Policy, practice, and prospects. In J. Horney (Ed.), *Criminal justice 2000* (Vol. 3, pp. 109-175). Washington, DC: U.S. Department of Justice.

Cullen, F. T., Van Voorhis, P., & Sundt, J. L. (1996). Prisons in crisis: The American experience. In R. Matthews & P. Francis (Eds.), *Prisons 2000: An international perspective on the current state and future of imprisonment* (pp. 21-52). New York: Macmillan.

Davies, W., & Burgess, P. W. (1988). The effect of management regime on disruptive behavior: An analysis within the British prison system. *Medical Science and Law*, *28*, 243-247.

*Davis, R. (1988). Education and the impact of the family reunion program in a maximum security prison. *Journal of Offender Counseling, Services, and Rehabilitation*, *12*, 153-159.

*Dietz, E. F., O'Connell, D. J., & Scarpitti, F. R. (2003). Therapeutic communities and prison management: An examination of the effects of operating an in-prison therapeutic community on levels of institutional disorder. *International Journal of Offender Therapy and Comparative Criminology*, *47*, 210-223.

DiIulio, J. J., Jr. (1987). *Governing prisons: A comparative study of correctional management*. New York: Free Press.

DiIulio, J. J., Jr. (1991). *No escape: The future of American corrections*. New York: Basic Books.

*Eisenberg, L., Lachman, R., & Moling, P. A. (1963). A psychopharmacologic experiment in a training school for delinquent boys: Methods, problems, findings. *American Journal of Orthopsychiatry*, *33*, 431-447.

*Empey, L., & Lubeck, S. (1971). *The Silverlake experiment*. Chicago: Aldine.

*Feder, B. (1962). Limited goals in short-term group psychotherapy with institutionalized delinquent adolescent boys. *International Journal of Group Psychotherapy*, *12*, 503-507.

Felson, M., & Clarke, R. V. (1998). Opportunity makes the thief: Practical theory for crime prevention. *Policing and Reducing Crime Unit* (Police Research Series Paper 98). London: PRC Unit Publication.

*Fox, V. (1954). The effect of counseling on adjustment in prison. *Social Forces*, *32*, 285-289.

*Friedland, D. M. (1960). *Group counseling as a factor in reducing runaway behavior from an open treatment institution for delinquent and pre-delinquent boys*. Unpublished doctoral dissertation, New York University, New York.

Gendreau, P. (1996). The principles of effective intervention with offenders. In A. T. Harland (Ed.), *Choosing correctional options that work* (pp. 117-130). London: Sage.

Gendreau, P. (2002). We must do a better job of cumulating knowledge. *Canadian Psychology*, *43*, 205-210.

Gendreau, P., Agee, V., & French, S. A. (2004). *Encouraging "good" behavior in prisons*. Unpublished manuscript, Centre for Criminal justice Studies, University of New Brunswick, Canada.

Gendreau, P., & Andrews, D. A. (2001). *Correctional Program Assessment Inventory—2000* (*CPAI-2000*). Saint John, Canada: University of New Brunswick.

Gendreau, P., & Goggin, C. (1997). Correctional treatment: Accomplishments and realities. In P. Van Voorhis, M. Braswell, & D. Lester (Eds.), *Correctional counselling and rehabilitation* (3rd ed., pp. 271-179). Cincinnati, OH: Anderson.

Gendreau, P., Goggin, C., & Cullen, F. (1999). *The effects of prison sentences on recidivism*. Report to the Corrections Research and Development and Aboriginal Policy Branch (Cat. No. J42-87/1999E), Solicitor General of Canada. Ottawa: Public Works and Government Services Canada.

Gendreau, P., Goggin, C., French, S., & Smith, P. (in press). Practicing psychology in correctional settings. In A. K. Hess & I. B. Weiner (Eds.), *The handbook of forensic psychology* (3rd ed.). New York: John Wiley.

Gendreau, P., Goggin, C., & Law, M. A. (1997). Predicting prison misconducts. *Criminal Justice and Behavior*, *24*, 414-431.

Gendreau, P., Goggin, C., & Smith, P. (2000). Generating rational correctional policies: An introduction to advances in cumulating knowledge. *Corrections Management Quarterly*, *4*, 52-60.

Gendreau, P., Goggin, C., & Smith, P. (2001). Implementation guidelines for correctional programs in the "real world." In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), *Offender rehabilitation in practice* (pp. 247-268). Chichester, UK: Wiley.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? *Criminal Justice and Behavior*, *29*, 397-426.

Gendreau, P., Goggin, C., & Smith, P. (2003). Erratum. *Criminal Justice and Behavior*, *30*, 722-724.

Gendreau, P., & Keyes, D. (2001). Making prisons safer and more humane environments. *Canadian Journal of Criminology*, *43*, 123-130.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, *34*, 575-607.

Gendreau, P., & Ross, R. R. (1979). Effective correctional treatment: Bibliotherapy for cynics. *Crime & Delinquency*, *25*, 463-489.

*Gendreau, P., Ross, R. R., & Izzo, R. (1985). Institutional misconduct: The effects of the UVIC program at Matsqui Penitentiary. *Canadian Journal of Criminology*, *27*, 209-217.

Gendreau, P., Tellier, M. C., & Wormith, J. S. (1985). Protective custody: The emerging crisis within our prisons. *Federal Probation*, *44*, 55-63.

*Gesch, C. B., Hammond, S. M., Hampson, S. E., & Eves, A. (2002). Influence of supplementary vitamins, minerals and essential fatty acids on the antisocial behavior of young prisoners: Randomized, placebo-controlled trial. *British Journal of Psychiatry*, *181*, 22-28.

*Goldenberg, F., & Cowden, J. F. (1977). An evaluation of intensive group psychotherapy with male offenders in isolation units. *Corrective and Social Psychiatry and Journal of Behavior Technology, Methods and Therapy*, *23*, 68-72.

*Goldstein, A. P., Glick, B., Reiner, S., Zimmerman, D., & Coultry, M. (1987). *Aggression replacement training: A comprehensive intervention for aggressive youth*. Champaign, IL: Research Press.

*Guerra, N. G., & Slaby, R. G. (1990). Cognitive mediators of aggression in adolescent offenders: 2. Intervention. *Developmental Psychology*, *26*, 269-277.

*Guttman, E. S. (1963). *MMPI measured changes in treated and untreated youth authority wards judged in need of psychiatric treatment* (Research Report No. 25). Sacramento: California Youth Authority.

Haertzen, C., Buxton, K., Covi, L., & Richards, H. (1993). Seasonal changes in rule infractions among prisoners: A preliminary test of the temperature-aggression hypothesis. *Psychological Reports*, *72*, 195-200.

Haney, C. (2003). Mental health issues in long-term solitary and “supermax” confinement. *Crime & Delinquency*, *49*, 124-156.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for analysis*. Toronto, Canada: Academic Press.

Hill, G. (1985). Predicting recidivism using institutional measures. In D. P. Farrington & R. Tarling (Eds.), *Prediction in criminology* (pp. 96-118). Albany: State University of New York Press.

*Hollin, C. R., & Courtney, S. A. (1983). A skills training approach to the reduction of institutional offending. *Personality and Individual Differences*, *4*, 257-264.

*Hollin, C. R., & Henderson, M. (1981). The effects of social skills training on incarcerated delinquent adolescents. *Journal of Behavioral Social Work and Abstracts*, *1*, 145-155.

Holt, R., & Phillips, R. (1991). Marion: Separating fact from fiction. *Federal Prisons Journal*, *2*, 28-36.

*Homant, R. J. (1976). Therapy effectiveness in a correctional institution. *Offender Rehabilitation*, *1*, 101-113.

Homant, R. J., & Witkowski, M. J. (2003). Prison deviance as a predictor of general deviance: Some correlational evidence from Project GANGMILL. *Journal of Gang Research*, *10*, 65-75.

Hunter, J. E., & Schmidt, F. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

*Ingram, G. L., Gerard, R., Quay, H. C., & Levinson, R. B. (1970). An experimental program for the psychopathic delinquent: Looking in the “correctional wastebasket.” *Journal of Research in Crime and Delinquency*, *7*, 24-30.

*Jacobson, F., & Magee, E. (1965). *Englewood project: Re-education: A radical correction of incarcerated delinquents*. Englewood, CO: Englewood Federal Youth Center.

*Johnson, B. R. (1987). Religiosity and institutional deviance: The impact of religious variables upon inmate adjustment. *Criminal Justice Review*, *12*, 21-30.

Johnson, W. W., Bennett, K., & Flanagan, T. J. (1997). Getting tough on prisoners: Results from the National Corrections Executive Survey, 1995. *Crime & Delinquency*, *43*, 24-41.

*Kassebaum, G., Ward, D. A., & Wilner, D. M. (1971). *Prison treatment and parole survival: An empirical assessment*. New York: John Wiley

*Kessemeier, L. A. (1966, March-April). Does group counselling pay its way? *Correctional Review*, pp. 26-30.

Keyes, D. (1996). *Preventing prison misconduct behavior: A quantitative review of the literature*. Unpublished honor's thesis, University of New Brunswick at Saint John, Canada.

*Kirigin, K. A., Braukmann, C. J., Atwater, J., & Wolf, M. M. (1982). An evaluation of teaching-family (Achievement Place) group homes for juvenile offenders. *Journal of Applied Behavior Analysis*, *15*, 1-16.

La Vigne, N. G. (1994). Rational choice and inmate disputes over phone use on Rikers Island. In R. V. Clarke (Ed.), *Crime prevention studies: Vol. 3* (pp. 109-125). Monsey, NY: Criminal Justice Press.

*Langenbach, M., North, M. Y., Aagaard, L., & Chown, W. (1990). Televised instruction in Oklahoma prisons: A study of recidivism and disciplinary actions. *Journal of Correctional Education*, *41*, 87-94.

Lawrence, S., & Travis, J. (2004). *The new landscape of imprisonment: Mapping America's prison expansion*. Washington, DC: Urban Institute, Justice Policy Center.

*Leak, G. K. (1980). Effects of highly structured versus nondirective group counseling approaches on personality and behavioral measures of adjustment in incarcerated felons. *Journal of Counseling Psychology*, *27*, 520-523.

*Leeman, L. W., Gibbs, J. C., & Fuller, D. (1993). Evaluation of a multi-component group treatment program for juvenile delinquents. *Aggressive Behavior*, *19*, 281-292.

*Levinson, R. B., & Kitchener, H. L. (1964). *Demonstration Counseling Project*. Washington, DC: National Training School for Boys.

*Liau, A. K. (1999). Evaluation of the peer helping component of a group treatment program for antisocial youth (Doctoral dissertation, Ohio State University, 1999). *Dissertation Abstracts International*, *60*(5-B), 2387.

Light, S. C. (1990). Measurement error in official statistics: Prison rule infraction data. *Federal Probation*, *54*, 63-68.

*Lipton, D., Martinson, R., & Wilks, J. (1975). *The effectiveness of correctional treatment: A survey of treatment evaluation studies*. New York: Praeger.

*Longhurst, J. E., & Mazer, G. E. (1988). The effects of a low glycemic diet on antisocial behavior in juvenile offenders. *International Journal of Biosocial Research*, *10*, 123-136.

Lovell, D., & Jemelka, R. (1996). When inmates misbehave: The costs of discipline. *Prison Journal*, *76*, 165-179.

Lowenkamp, C. T. (2004). *A program level analysis of the relationship between correctional program integrity and treatment effectiveness*. Unpublished doctoral dissertation, University of Cincinnati, Ohio.

*MacKenzie, D. L., & Shaw, J. (1990). Inmate adjustment and change during shock incarceration: The impact of correctional boot camps. *Justice Quarterly*, *7*, 125-150.

Matthews, B., Hubbard, D. J., & Latessa, E. (2001). Making the next step: Using evaluability assessment to improve correctional programming. *Prison Journal*, *81*, 454-473.

*McDougall, C. (1990). Anger control training with young offenders. In S. Duguid (Ed.), *The yearbook of correctional education* (pp. 319-324). Burnaby, Canada: Simon Fraser University.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*, 361-365.

*Medve, W. J. (1961). The rehabilitative aspects of team sports in a reformatory. *Journal of Correctional Education*, *13*, 4-11.

Morgan, R. D., & Flora, D. B. (2002). Group psychotherapy with incarcerated offenders: A research synthesis. *Group Dynamics*, *6*, 203-218.

*Morgan, R. D., Winterowd, C. L., & Fuqua, D. R. (1999). The efficacy of an integrated theoretical approach to group psychotherapy for male inmates. *Journal of Contemporary Psychotherapy*, *29*, 203-222.

*Morrissey, C. (1997). A multimodal approach to controlling inpatient assaultiveness among incarcerated juveniles. *Journal of Offender Rehabilitation*, *25*, 31-42.

*Moss, C. S., Hosford, R. E., Anderson, W. R., & Petracca, M. (1977). Personality variables of Blacks participating in a prison riot. *Journal of Consulting and Clinical Psychology*, *45*, 505-512.

Nesovic, A. (2003). *Psychometric evaluation of the Correctional Program Assessment Inventory*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.

*New York City Department of Correction. (1960, January). *A preliminary evaluation of the relationship between group psychotherapy and the adjustment of adolescent inmates in a short-term penal institution* (Research report). New York: Diagnostic staff of Rikers Island Penitentiary.

*Newburger, H. M. (1952). *The effect of group therapy upon certain aspects of the behavior and attitudes of institutionalized delinquents*. Unpublished doctoral dissertation, New York University.

Nossiter, A. (1994, September 17). Making hard time harder: States cut jail TV and sports. *The New York Times*, pp. A1, A10.

O'Donnell, I., & Edgar, K. (1996). *The extent and dynamics of victimizations in prisons*. London: Oxford.

*Ollendick, T. H., & Hersen, M. (1979). Social skills training for juvenile delinquents. *Behavior Research and Therapy*, *17*, 547-554.

Porporino, F. J., & Dudley, K. (1984). *An analysis of the effects of overcrowding in Canadian penitentiaries* (Working paper No. 6). Ottawa: Ministry of the Solicitor General of Canada.

*Prendergast, M., Farabee, D., & Cartier, J. (2001). The impact of in-prison therapeutic community programs on prison management. *Journal of Offender Rehabilitation*, *32*, 63-78.

*Pugh, D. N. (1993). The effects of problem-solving ability and locus of control on prisoner adjustment. *International Journal of Offender Therapy and Comparative Criminology*, *37*, 163-176.

*Queralt, J. P., Caballero, J., Casals, A., Navarro, J. C., & Serra, S. (1997). The treatment of aggressive prisoners: A closed regime programme. In S. Redondo, V. Garrido, J. Perez, & J. Barberet (Eds.), *Advances in psychology and law: International contributions* (pp. 442-451). New York: Walter de Gruyter.

Ralph, P., & Marquart, J. W. (1991). Gang violence in Texas prisons. *Prison Journal*, *71*, 38-49.

*Roberts, R. E., Cheek, E. H., & Mumm, R. S. (1994). Group intervention and reading performance in a medium-security prison facility. *Journal of Offender Rehabilitation*, *20*, 97-116.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*, 59-82.

*Ross, R. R., & McKay, H. B. (1976). A study of institutional treatment programs. *International Journal of Offender Therapy and Comparative Criminology*, *20*, 165-173.

*Roth, L. H., Rosenberg, N., & Levinson, R. B. (1971). Prison adjustment of alcoholic felons. *Quarterly Journal of Studies on Alcohol*, *32*, 382-392.

*Rudoff, A. (1960). *The effect of treatment in incarcerated young adult delinquents as measured by disciplinary history*. Unpublished master's thesis, University of Southern California, Los Angeles.

*Sarason, I. G., & Ganzer, V. J. (1973). Modeling and group discussion in the rehabilitation of juvenile delinquents. *Journal of Consulting Psychology*, *20*, 442-449.

*Schlichter, K. J., & Horan, J. J. (1981). Effects of stress inoculation on the anger and aggression management skills of institutionalized juvenile delinquents. *Cognitive Therapy and Research*, *5*, 359-365.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training researchers. *Psychological Methods*, *1*, 115-129.

Schnur, A. C. (1949). Prison conduct and recidivism. *Journal of Criminal Law, Criminology, and Police Science*, *40*, 36-42.

*Schoenthaler, S. J. (1983a). Diet and crime: An empirical examination of the value of nutrition in the control and treatment of incarcerated juvenile offenders. *International Journal of Biosocial Research*, *4*, 25-39.

*Schoenthaler, S. J. (1983b). The northern California diet-behavior program: An empirical examination of 3,000 incarcerated juveniles in Stanislaus County Juvenile Hall. *International Journal of Biosocial Research*, *5*, 99-106.

*Seckel, J. P. (1965). *Experiments in group counselling at two youth authority institutions* (Research Report No. 46). Sacramento: California Youth Authority.

Senese, J. D. (1997). Evaluating jail reform: A comparative analysis of podular/direct and linear jail inmate interactions. *Journal of Criminal Justice*, *25*, 61-73.

*Snyder, R., & Sechrest, L. (1959). An experimental study of directive group therapy with defective delinquents. *American Journal of Mental Deficiency*, *64*, 117-123.

*Sowles, R. C., & Gill, J. H. (1970). Institutional and community adjustment of delinquents following counseling. *Journal of Counseling and Clinical Psychology*, *34*, 398-402.

*Stallone, T. M. (1993). The effects of psychodrama on inmates within a structured residential behavior modification program. *Journal of Group Psychotherapy, Psychodrama & Sociometry*, *46*, 24-32.

*Sultan, F. E., Long, G. T., Kiefer, S. A., Schrum, D. M., Selby, J. W., & Calhoun, L. G. (1984). The female offender's adjustment to prison life: A comparison of psychodidactic and traditional supportive approaches to treatment. *Journal of Offender Counseling Services and Rehabilitation*, *9*, 49-56.

Toch, H. (1988). Social climate and prison violence. In M. C. Braswell, R. H. Montgomery, &L. X. Lombardo (Eds.), *Prison violence in America* (pp. 345-352). Cincinnati, OH: Anderson.

U.S. Department of Justice. (2004). *Prisoners in 2003* (Bureau of Justice Statistics Bulletin NCJ-205335). Washington, DC: P. M. Harrison & A. J. Beck.

Van Voorhis, P. (1994). Measuring prison disciplinary problems: A multiple indicators approach to understanding prison adjustment. *Justice Quarterly*, *11*, 679-709.

Van Voorhis, P., & Brown, K. (1996). *Evaluability assessment: A tool for program development in corrections*. Monograph prepared for the National Institute of Corrections. Washington, DC: U.S. Department of Justice.

Walkey, F. H., & Gilmour, D. R. (1984). The relationship between interpersonal distance and violence in imprisoned offenders. *Criminal Justice and Behavior*, *11*, 331-340.

Walmsley, R. (2003). *World prison population list.* London: Home Office.

*Walrath, C. (2001). Evaluation of an inmate-run Alternatives to Violence Project: The impact of inmate-to-inmate intervention. *Journal of Interpersonal Violence*, *16*, 697-711.

*Walters, G. D. (1999). Short-term outcome of inmates participating in the Lifestyle Change program. *Criminal Justice and Behavior*, *26*, 322-337.

*Watt, B. D., & Howells, K. (1999). Skills training for aggression control: Evaluation of an anger management programme for violent offenders. *Legal and Criminological Psychology*, *4*, 285-300.

Wener, R., & Olsen, R. (1978). *User based assessment of the federal metropolitan correctional centers—final report* (NCJRS No. 74794). Washington, DC: U.S. Federal Bureau of Prisons.

Williams, J. L., Rodeheaver, D. G., & Huggins, D. W. (1999). A comparative evaluation of a new generation jail. *American Journal of Criminal Justice*, *23*, 223-246.

*Wolk, R. L. (1966). The relationship of group psychotherapy to institutional adjustment. *International Journal of Group Psychotherapy*, *16*, 141-144.

*Wormith, J. S. (1984). Attitude and behavior change of correctional clientele: A three-year follow-up. *Criminology*, *22*, 595-618.

Wortley, R. (2002). *Situational prison control: Crime prevention in correctional institutions*. New York: Cambridge.

Wright, K. N. (2000). The evolution of decisionmaking among prison executives, 1975-2000. In J. Horney (Ed.), *Criminal Justice 2000* (Vol. 3, pp. 177-224). Washington, DC: U.S. Department of Justice.

Yeaton, W. H., & Wortman, P. M. (1993). On the reliability of meta-analytic reviews: The role of intercoder agreement. *Evaluation Review*, *17*, 292-309.

Zamble, E., & Porporino, F. J. (1988). *Coping, behavior, and adaptation in prison inmates*. New York: Springer-Verlag.

*Zivan, M. (1966). *Youth in trouble: A vocational approach* (Final report of a research and demonstration project, May 31, 1961-August 31, 1966). Dobbs Ferry, NY: Children's Village.