PERGAMON

Evaluation and Program Planning 23 (2000) 1–12

EVALUATION and PROGRAM PLANNING

www.elsevier.com/locate/evalprogplan

Wilderness challenge programs for delinquent youth: a meta-analysis of outcome evaluations

Sandra Jo Wilson*, Mark W. Lipsey

Vanderbilt University, Vanderbilt Institute for Public Policy Studies, 1207 18th Ave. So., Nashville, TN 37212, USA

Received 1 July 1998; received in revised form 1 February 1999; accepted 1 June 1999

Abstract

Evaluations of wilderness challenge programs were meta-analyzed to assess the impact on delinquent behavior. The overall mean effect size for delinquency outcomes was 0.18 ($N = 22$), equivalent to a recidivism rate of 29% for program participants vs 37% for comparison subjects. Program length was not related to outcome among short-term programs (up to 6 weeks) but extended programs (over 10 weeks) showed smaller effects overall. However, the most influential program characteristics were the intensity of the physical activities and whether the program included a distinct therapeutic component. Programs involving relatively intense activities or with therapeutic enhancements produced the greatest reductions in delinquent behavior.

1. Introduction

Wilderness challenge programs have received considerable attention as rehabilitative and preventive interventions for youth with behavior problems, especially juvenile delinquents (Wilson & Howell, 1993). In a wilderness challenge program, youth participate in a series of physically challenging activities, such as backpacking or rock climbing, usually in an outdoor environment. These programs vary widely in terms of settings, types of physical activities, and therapeutic goals, but their treatment concepts are grounded in the field of experiential education (Gass, 1993). Experiential education is based on the idea of "learning by doing" and involves learning opportunities that include direct experience intended to facilitate personal growth. Wilderness programs exemplify this "learning by doing" idea in a straightforward form: participants are immersed in activities that directly challenge their skills and self-concepts.

Wilderness programs seek to ameliorate antisocial behavior through two interrelated dimensions of experiential learning. First, by mastering a series of incrementally challenging physical activities, the participant experiences a pattern of success and thereby builds confidence, self-esteem, and a more internalized locus of control. This newly empowered individual, then, is presumably less likely to continue with a pattern of inappropriate or illegal behavior (Gass, 1993). The physical activities included in wilderness programs have several characteristic elements that facilitate this process. The activities are unfamiliar and demanding, the problems they present are concrete and incremental, stress is used constructively, and the consequences of failure are obvious and can be substantial (Kimball & Bacon, 1993).

The second dimension of experiential learning is related to the group orientation of wilderness programs (Wichmann, 1993). By solving challenging problems in situations that require positive group interaction and cooperation, participants are expected to learn prosocial interpersonal skills that will transfer to situations outside the program. Although some programs include solo activities aimed at building self-esteem, most of the challenges presented to the participants of wilderness programs are designed so that they cannot be

* Tel.: +1-615-343-7215; fax: +1-615-322-8081.
E-mail address: sandra.j.wilson@vanderbilt.edu (S.J. Wilson).

0149-7189/00/$ - see front matter
PII: S0149-7189(99)00040-3

managed by an individual acting alone. Success at rock climbing, river rafting, and even preparing food or shelter in primitive conditions requires communication and cooperation, which are intended to develop participants' social and interpersonal skills.

In attempting to rehabilitate problem youth by focusing treatment on physical challenge and interpersonal interactions, the proponents of wilderness programs assume that an external locus of control, low self-esteem, and poor interpersonal skills are key risk factors for delinquency. Amelioration of these factors, in turn, is presumed to lead to reductions in delinquency. Recent longitudinal research (e.g., Hawkins & Catalano, 1992; Huizinga, Loeber, & Thornberry, 1992) documents the correlation between psychological and interpersonal adjustment and delinquent behavior, though the causal relationship is not yet established. A recent review (Hawkins et al., 1998) and a synthesis of longitudinal studies (Lipsey & Derzon, 1998) provide further evidence for the association between psychological and interpersonal factors and later delinquency.

This study reports the results of a meta-analysis of outcome evaluations conducted on wilderness challenge programs for delinquent youth. Two other quantitative syntheses published recently offered evidence of the beneficial effects of wilderness challenge programs (Cason & Gillis, 1994; Hattie, Marsh, Neill & Richards, 1997). However, both of these meta-analyses focused primarily on non-behavioral outcomes, such as self-esteem and locus of control, and included all types of subjects (adults and adolescents, nonproblem and problem youth). Although this broad scope allowed those authors to examine a wider range of research than we have covered here, our primary purpose is to address the specific impact of wilderness challenge programs on juvenile delinquency.

In addition, the two previous meta-analyses relied extensively on studies of the difference between a single treatment group before and after participation in a wilderness program rather than differences between a treatment and a comparison group. Tracking individuals over time without the benefit of a comparison group leaves considerable uncertainty about whether the observed effects are due to the intervention of interest or some other factor. The meta-analysis presented here, therefore, included only studies that employed a comparison group in an evaluation of the effectiveness of a wilderness challenge program for problem youth.

2. Method

Studies were selected for this meta-analysis based on a set of detailed criteria developed from the program theory outlined above, Lipsey's (1992) meta-analysis of delinquency treatment, and a review of a sample of wilderness program evaluation studies. An attempt was made to define selection criteria that would assure some degree of methodological rigor in the eligible studies without being so restrictive that they would exclude a significant proportion of the literature. These eligibility criteria were as follows:

1. The study had to involve wilderness challenge programs to reduce or prevent antisocial behavior or delinquency using treatment that included both a physical challenge element and an interpersonal element. Boot camps, recreational programs (e.g., midnight basketball, after-school sports, etc.), and other such programs involving physical activity were not eligible unless they specifically incorporated both the challenge and the interpersonal dimensions in planned activities. In addition, wilderness programs targeted exclusively on substance abuse and no other components of antisocial behavior or delinquency were not eligible.
2. The program had to be provided to antisocial or delinquent youth between the ages of 10 and 21 or directed toward changing antisocial or delinquent behavior as outcome variables. Thus, studies in which general youth samples were asked to self-report delinquent activities were considered eligible, but studies in which such youth completed only self-esteem measures were not eligible. On the other hand, studies that involved samples of antisocial youth but did not use behavioral outcomes were eligible (e.g., studies in which only the self-esteem of antisocial or delinquent youth was measured).
3. Only studies using a control or comparison group design were eligible. Comparison groups could be either randomized or nonrandomized but, if nonrandomized, had to utilize a matched comparison group or provide some evidence regarding pretest equivalence between the treatment and comparison groups. A control could be a placebo, wait-list, no treatment, or "treatment as usual" group. One group pretest–posttest studies were not eligible.
4. Results on outcome measures had to be presented in quantitative form and permit computation or reasonable estimation of an effect size in the form of the standardized difference between means.
5. Finally, to ensure that the research was relatively modern and accessible to the authors, studies had to be reported after 1950 and in English.

2.1. Identification and retrieval of eligible research reports

An attempt was made to identify and retrieve the entire population of empirical studies on wilderness

challenge programs that met the eligibility criteria, including both published and unpublished studies. Three sources were used to identify potentially eligible research reports. First, the bibliographies of previous literature reviews and meta-analyses (e.g., Lipsey, 1992) were reviewed. Second, a comprehensive search of bibliographic databases, including Psychological Abstracts, Dissertation Abstracts International, Educational Resources Information Center (ERIC), US Government Printing Office publications, National Criminal Justice Reference Service, and others identified a number of possible reports. Finally, the bibliographies of retrieved studies were themselves examined for related studies. Potentially eligible studies were found in the library, obtained via interlibrary loan, or requested directly from the author(s). We obtained more than 95% of the reports identified as potentially eligible and believe that there are very few eligible studies that remain unretrieved or unidentified.

2.2. Coding of the research reports

All eligible research reports were coded by the first author using a detailed computerized coding scheme based on the coding manual used in Lipsey's (1992) meta-analysis, the program theory outlined above, and examination of a number of wilderness program evaluations. The author coded each report twice, with an interval of no less than several days between them, to ensure that coding was consistent. The two coding passes yielded only trivial differences between them, which were reconciled by reviewing the report further to select the best alternative. Two general categories of information were coded, effect size information and study descriptor information.

2.3. Effect size coding

For each study, the quantitative outcome variables that contrasted the treatment and comparison groups were identified. These outcome variables were divided into two general groupings comprising those that assessed antisocial or delinquent behavior and those that represented interpersonal and psychological characteristics associated with delinquent or problem behavior (e.g., self-esteem, locus of control, school performance, etc.). An effect size estimate was computed for each relevant outcome variable for which sufficient quantitative information was reported. The index used for this purpose was the standardized difference between means, defined as the difference between the mean score for the treatment group and that of the comparison group divided by the pooled standard deviation of those scores. Pretest and follow-up effect sizes (i.e., from additional measurement waves) were also coded when possible. The computation of effect sizes was completed with the aid of Lipsey and Wilson (1996).

When possible, the posttest effect sizes were adjusted for pretest differences between the experimental and comparison groups. This adjustment was obtained in one of two ways — either the primary study presented pretest-adjusted means (or comparable statistics), or the coder adjusted the posttest effect size for pre-intervention differences by subtracting the pretest effect size from the posttest effect size.

2.4. Coding of study descriptors

In addition to effect size results, it is important to examine the relationships between the effects and various study characteristics, such as the methodology used, the nature of the treatment, and so forth. Thus, along with the effect size coding, several other categories of variables were coded.

2.4.1. Method variables

A wide range of information pertaining to study design and method was coded, including characteristics of the experimental design, measures, samples, attrition, and initial equivalence between treatment and comparison groups.

2.4.2. Study context

This category of coding included the year and form of publication.

2.4.3. Treatment

Specific information about the nature of the intervention was coded, including duration and intensity of treatment, physical setting of the program, the presence/absence of therapeutic enhancements, and the like. This information was coded in order to help identify the characteristics of the most and least effective programs.

2.4.4. Participants

Finally, study outcomes may vary with the type of juveniles treated in the programs. For example, programs may be more effective with younger juveniles. Where possible, information was coded about the age, sex, ethnicity, and prior delinquency history of the program recipients.

3. Results

The search of the literature yielded 28 eligible research studies, involving over 3000 individuals. Table 1 shows that the typical study was unpublished, had a sample size of more than 50 juveniles, and used a nonrandom method to assign subjects to treatment and

Table 1
General characteristics of the studies included in the meta-analysis

	N	%[a]
Study characteristics		
Type of publication		
Published (journal article, book chapter)	10	36
Unpublished (dissertation, technical report)	18	64
Sample size (post treatment)		
≤ 50	8	29
50–100	12	43
> 100	8	29
Assignment method		
Random	12	43
Nonrandom	16	57
What control or comparison group received		
No treatment or wait list control	8	29
Probation	8	29
Institutionalization	10	36
Other	2	7
Subject characteristics		
Age		
10–12 years	3	11
13–15 years	17	61
16–18 years	8	29
Ethnicity		
Anglo/Caucasian (groups are >90% Anglo)	16	57
Mixed race (groups include minority groups and Anglos)	4	14
Unknown (not reported)	8	29
Level of delinquency risk		
Non-delinquent/at-risk	4	14
Delinquent: probation or adjudication	16	57
Institutionalized	7	25
Mixed (full range of risk)	1	4
Treatment characteristics		
Therapeutic enhancements		
No specific therapy	18	64
Additional therapeutic activity/treatment	10	36
Rated intensity of challenge activities		
Low	3	11
Medium	9	32
High	16	57
Duration of treatment (weeks)		
≤ 1 week	9	32
3–6 weeks	13	46
> 10 weeks	5	18
Unknown	1	4

[a] Some percentages may not add up to 100 because of rounding.

comparison groups (e.g., matching based on personal or demographic characteristics). The subjects in most of the studies were male Caucasians, between the ages of 13 and 15, who were on probation and/or adjudicated delinquents. In addition, most of the wilderness challenge programs were rated by the coder as having medium to high intensity challenge activities, were longer than two weeks in length, and had no therapeutic component beyond the standard wilderness program treatment elements.

In ten of the studies, the intervention and comparison groups were institutionalized, which raises a question about the opportunity these children had to commit new offenses subsequent to the program period. Spurious group differences would result if outcomes were measured on wilderness program participants who returned to the community after completing their program while their control counterparts remained institutionalized. In all ten of these studies, however, either the treatment and control subjects both remained insitutionalized and antisocial behaviors were measured in the institution, or treatment and control subjects both returned to the community and outcomes were measured in the community.

3.1. Effect sizes for delinquency and non-delinquency outcomes

The 28 studies yielded 60 effect sizes on outcome variables which are summarized in Table 2. For studies that contributed more than one effect size for a given outcome construct, the mean of those effect sizes was used. This procedure ensured that each study sample contributed only one effect size to the total distribution of effects on each outcome construct. Effect sizes for pretest and follow-up comparisons between treatment and comparison groups were also computed where possible but there were too few estimates for further analysis. The second column of Table 2 shows the mean effect size for each outcome construct. The effect sizes were weighted by the inverse variance when calculating these means to account for the different sample sizes on which each effect size was based (Shadish & Haddock, 1994).

The antisocial behavior and delinquency construct included antisocial behavior, officially recorded recidivism, and self-reported delinquent behavior. Note that Table 2 shows two mean effect sizes for antisocial behavior outcomes. The distribution of effect size estimates in this category had two large negative outliers (−0.85 and −0.42). We were reluctant to remove these two studies from the analysis because we have so few studies to begin with and because removing them may lead to an overestimate of the effectiveness of wilderness challenge programs. On the other hand, they are clearly atypical values. An examination of the two studies that produced the outliers revealed nothing especially unusual about them; the study, subject, and method characteristics were all within the range of characteristics exhibited by the other studies in the sample. Mean effect size values are therefore presented with and without these outliers.

The overall mean effect size of 0.18 for antisocial behavior and delinquency outcomes was positive, though somewhat modest, indicating that, on average, treatment groups exhibited less antisocial and delinquent behavior than comparison groups after treatment. Moreover, the 95% confidence interval did not

Table 2
Mean posttest effect size for each outcome construct

Outcome construct	Weighted mean ES	*N*	95% C.I.		*Q*
			Lower	Upper	
Antisocial behavior and delinquency					
All effect sizes	0.18	22	0.10	0.27	42.17*
Less outliers	0.24	20	0.15	0.32	25.23
Interpersonal adjustment (social skills)	0.28	7	0.06	0.49	12.51
Locus of control[a]	0.10	7	−0.10	0.30	16.48*
Self-esteem[a]	0.31	9	0.14	0.47	22.53*
Other psychological adjustment (e.g., clinical measures)	0.25	7	0.09	0.42	9.03
School adjustment	0.30	6	0.11	0.49	7.61

* $P < 0.05$.
[a] One effect size outlier was excluded from the mean calculations.

include zero, indicating that the mean was statistically significant.

The modal outcome measure for delinquent behavior in these studies was rearrest recidivism and the majority of program participants were identified as officially delinquent at the time of intervention. Thus, it is informative to convert the 0.18 effect size from standard deviation units to the equivalent recidivism rates. We have done this using the arcsine transformation (Cohen, 1988) and the mean comparison group recidivism rate derived from our sample. Ten of the 22 studies contributing delinquency or antisocial behavior effect sizes used recidivism as their outcome measure. The average control group recidivism rate from the 10 studies using recidivism as an outcome was 37%. The 0.18 mean effect size translates to about an 8% point drop from a 37% baseline; that is, about a 29% recidivism rate for the treatment group. To put the mean effect size in this more intuitive form, therefore, on average only about 29% of the wilderness program participants will recidivate compared to 37% of the control participants. This reduction, though modest, is not trivial, and could represent a considerable number of juveniles.

The other outcome categories listed in Table 2 are non-behavioral outcomes and cover general interpersonal and psychological variables thought to be related to antisocial and delinquent behavior. For example, because wilderness programs attempt to improve youths' confidence in themselves, an assessment of their impact on self-esteem is relevant. The mean effect size values were positive for all the interpersonal and psychological adjustment constructs, indicating that juveniles in wilderness programs, on average, showed better outcomes on these variables than the control youth. In addition, the 95% confidence intervals for the mean effect sizes on all outcomes except locus of control did not include zero, indicating that they were statistically significant. Note, however, the small number of studies contributing to the mean effect in each of these categories.

The last column in Table 2 presents the results of the homogeneity tests. The *Q* statistic tests whether the observed variability in the distribution of effect size estimates is greater than would be expected from sampling error. The significant *Q* for the antisocial behavior outcomes (when the outliers were included), as well as for self-esteem and locus of control, implies that the mean effect size in these instances may not be an accurate representation of the outcomes of the studies in the sample. Without the outlying cases, however, the effect size distribution for the delinquency outcomes was homogeneous (i.e., the *Q* statistic was not significant). Despite this finding, we believed that there was sufficient variability to warrant further analysis because the confidence interval was relatively large (0.15–0.32) and because the small number of effect sizes resulted in a *Q* test with relatively little power for rejecting the null hypothesis of homogeneity.

Additional analysis was therefore applied to examine the observed variability in effect sizes and its relationship to differences in the methodology and differences in the treatments and subjects used in the various evaluation studies. It is to this topic that we now turn.

3.2. Analyzing effect size variability

The analyses presented here focus on the antisocial behavior and delinquency outcomes. Further analysis of the other outcome constructs was not possible because of the small number of effect sizes in each of those categories. We analyzed the delinquency outcome data with and without the outliers to determine whether the outliers caused the results to differ. The same pattern of results emerged in both analyses, therefore we present below only the analyses that included the two outliers.

Table 3
Weighted correlations among antisocial behavior effect sizes and study descriptors

	Method cluster				Subject cluster		Treatment cluster		
Variable	Assignment method	Type of outcome measure	Pretest adjustment	Group similarity	Delinquency risk	Age	Duration	Intensity	Therapy
Effect size	0.12	−0.35	−0.11	−0.04	−0.38	−0.29	−0.36	0.14	−0.42*
Assignment method		−0.06	0.12	0.46*	0.21	−0.45*	0.05	0.01	−0.47*
Type of outcome measure			0.32	−0.05	0.40	0.29	0.34	0.36	0.50*
Pretest adjustment to ES				−0.08	0.06	0.22	0.29	−0.03	0.26
Group similarity at pretest					0.51*	−0.36	−0.09	−0.31	−0.22
Delinquency risk						−0.12	−0.03	−0.08	0.42*
Age							0.55*	−0.07	0.18
Duration of treatment								0.33	−0.12
Intensity of treatment									0.03

* $P < 0.05$ (two-tailed).

To analyze the variability of the effect sizes as a function of study characteristics, a weighted multiple regression approach was used in which the weights were the inverse variance of each effect size (Hedges & Olkin, 1985). Table 3 presents the matrix of weighted correlations for effect size and study characteristics.

The first row of the matrix reports the weighted correlations between the study characteristics and effect size. The effect sizes represent the antisocial behavior outcomes and are the dependent variable in this analysis; the study characteristics are the independent variables. The remaining rows in the matrix show the intercorrelations among the independent variables classified into three clusters: a method cluster, a subject characteristics cluster, and a treatment cluster.

3.2.1. Method cluster

The method cluster contains four variables that describe study methodology. Assignment method is the way subjects are assigned to treatment and comparison groups (0 is random; 1 is nonrandom). Type of outcome measure describes how outcome information was obtained (0 is self-report surveys; 1 is records, archives). The pretest adjustment variable indicates whether the effect size estimate was adjusted for pretest differences between treatment and comparison groups (1 is adjusted; 2 is not adjusted). Group similarity is the coder's assessment of the pre-intervention equivalence between the treatment and comparison groups given the information provided in the study report. This rating was made on a scale of one to seven, with a rating of one indicating high similarity.

3.2.2. Subject cluster

The cluster of study characteristics related to the research subjects contains two variables. Delinquency risk indicates how involved the study samples were in delinquency and ranges from nondelinquent, but judged at-risk, to institutionalized offenders (see Table 1). Age was recorded as the mean for the subjects in the study sample. Analysis of the role of the gender and ethnic mix of the study samples would have been informative but, unfortunately, only two studies included more than a few females and only three included more than a few minority group members. This precluded any examination of the relationships between program effectiveness and gender or ethnicity.

3.2.3. Treatment cluster

The treatment cluster contains three variables: program duration, program intensity, and therapy. Duration was coded as the actual number of weeks spent in the program with missing values set to the median (4 weeks). However, the programs all fell into three distinct groups: (a) less than one week (usually weekends), (b) the most typical length, 3–6 weeks, and (c)

longer duration programs that generally lasted 3 months or more. To reflect this natural clustering in the analysis, duration was recoded into these three categories. Program intensity, in turn, was described by a rating of the physical rigor and difficulty of the challenging activities (1 is low intensity; 7 is high intensity). For example, programs that involved whitewater rafting, wagon trains, and solo and/or group backpacking trips were rated high intensity programs. Programs with high ropes courses, day hikes, indoor climbing, cabin camping, etc. were rated as mid-range intensity programs. Programs with nothing more challenging than low ropes courses, "trust falls," and the like were rated as low intensity programs.

The therapy component of these programs was coded with a dichotomous variable that differentiated those with a distinct therapeutic component from those without one (0 is therapy; 1 is no therapy). Some wilderness programs relied entirely on the challenge and associated interpersonal interaction as the rehabilitative elements of the program and did not include other more directly therapeutic components. Other programs incorporated an explicit therapeutic component. The therapeutic wilderness programs in our sample employed a wide variety of techniques including transactional analysis, behavior management, family therapy, reality therapy, cognitive/behavioral techniques, and therapeutic group process sessions in the evenings. In general, the therapeutic component was intended to reduce antisocial behavior and help the participants apply what they were learning during the program to outside contexts and settings.

3.3. Relationships of study characteristics to effect size

The first row of the matrix in Table 3 shows that the therapy variable in the treatment cluster is significantly correlated with effect size and that a number of the other study characteristics have relatively large correlations that fall short of statistical significance. In addition, a number of the intercorrelations between different study characteristics in the remaining rows were significant and large enough to raise a question about the statistical independence of these characteristics. For instance, what appear to be relationships between treatment characteristics and effect size may only reflect differences in study method that are confounded with the nature of the treatment. Multiple regression analysis was therefore used in an attempt to better identify the independent components of the relationships between study characteristics and effect size.

The initial multiple regression results are presented in Table 4. As noted earlier, this analysis includes the two effect size outliers and is conducted as a weighted analysis using the inverse variance weight for each effect size. The variables included in the model presented in Table 4 account for a significant proportion of the observed variability in the effect size estimates ($R^2 = 0.58$, $P < 0.05$) and leave a nonsignificant residual. In spite of the relatively high zero-order correlations of some of the method and subject variables with effect size, only the pretest adjustment dummy code and the three treatment variables (duration, intensity, and therapy) made individually significant independent contributions to the prediction of effect size. This finding indicates that a more parsimonious model is warranted.

A reduced multiple regression model is therefore presented in Table 5. When the nonsignificant predictor variables were removed from the model, only the three treatment variables continued to be significant; the previously significant pretest adjustment variable was not significant in the reduced model. The three treatment variables in the reduced model accounted for a significant proportion of the observed variability in effect size estimates ($R^2 = 0.43$, $P < 0.05$). Also, as before, the residual was not significant, indicating that the variance unaccounted for after fitting the re-

Table 4
Results of the initial weighted multiple regression analysis[a]

	Beta	Q
Method cluster		
Assignment method	−0.39	2.49
Type of outcome measure	−0.07	0.10
Pretest adjustment	0.50	4.25*
Group similarity rating	0.18	0.55
Subject cluster		
Delinquency risk	−0.14	0.31
Age	0.08	0.09
Treatment cluster		
Duration	−0.71	8.31*
Intensity	0.53	5.53*
Therapy	−0.59	5.32*

* $P < 0.05$.
[a] Model: Q (9) = 24.46, $P < 0.05$; Residual: Q (12) = 17.71, ns; $R^2 = 0.58$.

Table 5
Results of the reduced weighted multiple regression analysis[a]

	Beta	Q
Treatment cluster		
Duration	−0.35	9.96*
Intensity	0.14	3.96*
Therapy	−0.42	9.77*

* $P < 0.05$.
[a] Model: Q (3) = 18.22, $P < 0.05$; Residual: Q (18) = 23.95, ns; $R^2 = 0.43$.

gression model was not greater than expected from sampling error.

The significant contribution of the intensity variable to the final model shows that high intensity wilderness challenge programs, such as those that employ strenuous solo and group expeditions and other difficult physical activities, produced larger delinquency reductions than programs that employ less rigorous activities. In addition, the therapy variable made an independent contribution above and beyond the effects associated with intensity. The challenge programs that incorporated a distinct therapy component resulted in lower delinquent and antisocial behavior than those without such enhancements.

The relationship of the duration variable to effect size was peculiar and requires some explanation. The direction of this relationship was negative, indicating that longer duration programs were associated with smaller effects on antisocial behavior and delinquency outcomes than shorter programs. That is, youth who participated in longer programs showed less delinquency reduction relative to controls than those who attended short-term programs. This result is counterintuitive and it may not be appropriate to interpret it exclusively in terms of program duration; that is, longer programs may have some other associated characteristics that we were unable to describe in our coding that account for this effect.

In order to examine the duration variable in more detail, we fit the regression model shown in Table 5 to only the short- and medium-term programs ($N = 17$). Actual program duration was represented in weeks rather than with the original three categories (a range of less than 1–6 weeks). Program duration for the short- and medium-term programs was not significantly related to effect size. Thus, it seems likely that the long-term programs are somehow categorically different from the more common shorter programs in some way that results in smaller effects on delinquency outcomes. Examination of the study reports revealed several features of the five long-term programs that might explain their counterintuitive results. First, all five of the programs in this category had multiple treatment elements such that the wilderness challenge program was only a portion of the total treatment received by the participants. The balance of the time was spent in institutional settings or other non-challenge type activities such as community service. The challenge portion of the programming, therefore, may not have had the long duration indicated by total program length. Also, either the implementation or the effects of the challenge portion may have been diluted by being combined with the other elements of these combination programs. It might be the case, as well, that these programs are not as effective in institutional settings or when applied to institutionalized offenders. Unfortunately, insufficient information was available in the study reports to permit direct examination of any of these hypotheses.

An additional consideration is the possibility that the duration and therapy variables interact in a way that the regression analysis was not designed to reveal. Table 6 presents the mean effect size for delinquency outcomes broken out by therapy and duration (short- and medium-term programs vs long-term programs). The pattern of these means indicates that absence of therapy in the long-term programs may be especially detrimental. The two studies of long-term programs without therapy reported an average effect size that was negative while those with therapy showed a positive effect. Conversely, the presence of therapy in the shorter term programs appears to be especially effective. The mean effect size for that combination is more than double that for any other configuration in Table 6.

Finally, in order to give the reader a better idea of the characteristics of the treatments, subjects, and methods in each contributing primary study, we present a brief description of each program in Table 7.

4. Discussion

Are wilderness challenge programs effective for reducing antisocial and delinquent behavior? The results presented above lead us to answer with a qualified yes. We suggest, however, that the results of this meta-analysis be interpreted chiefly in a formative manner. Our results identify the wilderness program variants that seem to be working best given current evidence, but say even more about what must still be learned before it will be possible to draw convincing conclusions about the effectiveness of wilderness programs and how best to optimize them for reducing antisocial and delinquent behavior.

The moderately positive overall results found for the wilderness challenge programs with outcome evaluations, nonetheless, support the value of such programs as an intervention for delinquent youth. Programs involving relatively intense physical activities and therapeutic enhancements such as individual counseling, family therapy, and therapeutic group

Table 6
Mean effect sizes for delinquency outcomes: therapy by duration

Group	Therapy		No therapy		Total	
	ES	(*N*)	ES	(*N*)	ES	*N*
1–6 week programs	0.51	(4)	0.22	(13)	0.30	(17)
Long-term programs	0.20	(3)	−0.14	(2)	0.00	(5)
Total	0.36	(7)	0.11	(15)	0.18	(22)

Table 7
Program, subject, and design characteristics of contributing studies

Study	Program	Sample	Design	Deliquency effect size
Andrew (1977)	A low intensity 5-day wilderness program with no additional therapeutic enhancement.	30 delinquent boys and girls (mean age = 13.7) received either the wilderness program or probation.	Non-random	−0.85
Birkenmayer and Polonoski (1976)	A high intensity 13-week wilderness program with no additional therapeutic component.	148 institutionalized, male adolescents (mean age = 15) were placed in the wilderness program or remained in the training school.	Non-random	−0.42
Burdsal and Buel (1980)	A high intensity 6-week therapeutic camping program including the following therapeutic techniques: transactional analysis, behavior management, and counseling with parents.	173 11-year old boys and girls who were at risk for later delinquency received either the wilderness program or no treatment.	Non-random	0.65
Callahan (1989)	A very high intensity 25-week wilderness program that included behavioral, cognitive-behavioral, and group problem solving therapies, and parent counseling.	70 delinquent boys and girls (mean age = 15) received either the wilderness program or probation.	Non-random	0.20
Carter (1979)	A high intensity 4-week wilderness program with no additional therapeutic component.	120 institutionalized males (mean age = 13.3) received either the wilderness program or treatment as usual in the institutional setting.	Non-random	0.26
Castellano and Soderstrom (1992)	A very high intensity 4-week wilderness program that included reality therapy and group problem solving.	56 delinquent males (mean age = 15.5) were assigned to the wilderness program or placed on probation.	Non-random	0.41
Chiles (1984)	A very high intensity 3.5-week wilderness program that included no specific therapeutic component.	56 institutionalized males (mean age = 16.2) received either the wilderness program or treatment as usual in the institutional setting.	Non-random	0.15
Elrod and Minor (1992)	A moderate intensity 3-day outdoor adventure program with no additional therapeutic component.	43 male and female probationers (mean age = 14.8) received either the outdoor adventure program or regular probation services.	Random	0.09
Greenwood and Turner (1987)	A very high intensity 56-week wilderness program with group therapy and family therapy.	266 seriously delinquent males (mean age = 16.3) were assigned to the wilderness program or an institution.	Non-random	0.14
Hileman (1979)	A very high intensity 4-week wilderness program with no additional therapeutic component.	96 male delinquents (mean age = 14.9) were assigned to the wilderness program or probation.	Non-random	0.16
Kelly and Baer (1971)	A very high intensity 4-week wilderness program that used no specific therapy in conjunction with the wilderness experience.	120 15–17 year old male delinquents were assigned to the wilderness program or a training school.	Non-random	0.48
Minor (1988)	A moderate intensity 3-day outdoor adventure program with no additional therapeutic enhancement.	40 male and female probationers (mean age = 15.4) received either the outdoor adventure or regular probation services.	Random	−0.07
Molof (1967)	A high intensity 28-week wilderness camp that included no therapeutic component.	370 18-year old institutionalized males received either the wilderness camp or services as usual in an institution.	Random	−0.02
Ramirez (1976)	A very high intensity 3 week wilderness camp with no additional therapeutic enhancements.	86 delinquent males (mean age = 16.5) received either the wilderness camp or were placed on a waiting list.	Random	0.15
Ramirez (1976)	A very high intensity 3-week wilderness camp with no additional therapeutic enhancements.	145 male delinquents (mean age = 15.5) received either the wilderness camp or were placed on a waiting list.	Random	−0.02
Rohling (1987)	A very high intensity 4-week wilderness program that used no specific therapy in conjunction with the wilderness experience.	38 delinquent males and females (mean age = 14) were assigned to the wilderness program, an institution, or received services in the community.	Non-random	0.42
Sachs and Miller (1992)	A moderate intensity 3-day wilderness program that used no specific therapy in conjunction with the wilderness experience.	13 males and 3 females (mean age = 16) who were institutionalized received either the wilderness program or relaxation training.	Random	0.52

(continued on next page)

Table 7 (*continued*)

Study	Program	Sample	Design	Deliquency effect size
Weeks (1985)	A high intensity 21-week residential adventure probation program that included reality therapy and counseling with parents.	36 male and female delinquents (mean age = 15) were placed in the adventure program or on probation.	Non-random	0.65
Wichmann (1990)	A very high intensity 4-week therapeutic wilderness program that included reality therapy.	72 15-year old boys and girls who were delinquent or predelinquent were placed in either the wilderness program or received community-based advocacy services.	Random	0.55
Willman and Chun (1973)	A very high intensity 6-week wilderness program that used no specific therapy in conjunction with the wilderness experience.	253 16–18 year old male delinquents were assigned to the wilderness program or a training school.	Random	0.48
Winterdyk (1980)	A high intensity 3-week wilderness program that included no additional therapeutic enhancements.	58 male delinquents (mean age = 14.8) received either the wilderness program or probation.	Random	0.07
Ziven (1988)	A very low intensity 5-day challenge program that included group therapy.	73 institutionalized males and females (mean age = 15) were assigned to the challenge program or treatment as usual in the institutional setting.	Non-random	0.31

sessions appear to be especially effective. A typical program with these key features might include strenuous solo and group backpacking expeditions as well as nightly group therapy sessions and individual counseling for specific problem areas.

However, the results described here apply primarily to Caucasian boys who are already "in the system;" that is, boys who have been arrested and convicted of a crime. The existing evaluation literature does not provide sufficient instances of challenge programs applied to other types of participants to permit an examination of their effectiveness with girls, members of minority groups, or predelinquent youth. An important question for program evaluation in this area, therefore, is how well these programs work for other populations of juveniles who are at risk for delinquency or already delinquent.

Perhaps the most troubling area of uncertainty in this body of evaluation research has to do with the counterintuitive dose-response relationship that appears for program duration. One would expect that more exposure to and involvement in wilderness challenge programs, at least up to some reasonable level, would produce greater effects on delinquency. After that, the effect might plateau, but it does not seem plausible that it would dramatically decrease. Our analysis showed that program length was not related to the magnitude of the effect on delinquency among the short- and medium-term (less than 6 weeks) programs. It seems most likely that, above that point, the duration variable acted as a proxy for some other characteristics of extended programs that account for their diminished effectiveness. The information reported in the available studies, unfortunately, is not sufficient to identify any such characteristics and this topic clearly warrants further investigation.

We did observe, however, that most of the longer duration programs represented in the current evaluation literature incorporated other elements of treatment along with the wilderness challenge component. Proponents of wilderness programs often discuss the importance of having certain "defining" experiences that result from the challenges a participant must meet. These capstone experiences are presumed to trigger changes in a participant's self-esteem and antisocial behavior. Perhaps these defining moments are less likely or less effective when they are only a part of a youth's overall treatment program. Experimental or quasi-experimental outcome studies that compare programs with different combinations of program elements, such as wilderness programs vs wilderness programs with community service, are required to study this issue in more detail.

Whatever the role of duration, the results of our meta-analysis provide rather strong indications that a therapeutic component enhances the delinquency effects of challenge programs. Perhaps its role is to help the participating youth focus their attention on the defining experiences of the challenge program. That is, incorporating therapy into a wilderness challenge program may enable participants to process the challenging experiences and draw implications for their own behavior with more success. The larger effects of programs that involve more intense challenge activities is consistent with this interpretation. If the salience of the challenge and psychological engagement with it are critical to the positive changes induced by these programs, more intense challenges may facilitate these responses independent of any therapeutic facilitation.

The lack of sufficient evaluation studies to disentangle the influence of program duration and therapeutic enhancements upon program effects leads to another problem. It is of considerable practical importance to know if therapeutic enhancements are essential for optimal effects on delinquency and which therapy components are best. Moreover, it is intriguing that the largest effects found among existing evaluation studies are for relatively short-term programs with therapeutic enhancements. If longer duration is not necessary to produce optimal effects, very cost-effective program variants should be possible. The evaluation literature currently includes too few studies of challenge programs with different combinations of therapy components and program lengths to permit this issue to be resolved.

Finally, the finding that studies which employed more rigorous methodologies, such as random assignment to treatment and comparison groups, did not result in different observed effects gives some warrant for including them as useful evidence, an advantage since so few randomized studies have been conducted. This does not imply, however, that methodological quality does not make a difference regarding the validity of effect size estimates. Randomized studies still provide the most convincing evidence for a causal relationship between participation in wilderness programs and reductions in delinquent and antisocial behavior. Perhaps the most important finding of our meta-analysis is that additional randomized studies of wilderness challenge programs, especially those with therapeutic enhancements, are clearly justified by the promising findings of the studies reviewed here. A larger body of evidence will be necessary before it will be possible to generate conclusive results about the effects of these programs and the influence of study methodology on the size of the effects actually observed using different methods.

References[1]

Andrew, S. H. (1977). *An evaluation of two stress-challenge programs for delinquent youth*. Unpublished master's thesis, Southern Illinois University, Carbondale*.

Birkenmayer, A. C., & Polonoski, M. (1976). *The community adjustment of male training school recidivists: Part II, The D.A.R.E. experience*. Ontario, Canada: Ministry of Correctional Services, Planning and Support Services Division *.

Burdsal, C., & Buel, C. L. (1980). A short term community based early stage intervention program for behavior problem youth. *Journal of Clinical Psychology*, *36,* 226–241 *.

Callahan, R. C. (1989). *Academic and therapeutic potential of the Sierra II process: An evaluation of an adapted Outward Bound diversion program for adjudicated juvenile delinquents*. Unpublished doctoral dissertation, Old Dominion University, Norfolk, Virginia*.

Carter, J. D. (1979). *A comparative study of two training programs in Florida for delinquent youths: Eckerd Wilderness Camping and State training schools*. Pensacola, FL: University of West Florida, Educational Research and Development Center *.

Cason, D., & Gillis, H. L. (1994). A meta-analysis of outdoor adventure programming with adolescents. *Journal of Experiential Education*, *17,* 40–47.

Castellano, T. C., & Soderstrom, I. R. (1992). Therapeutic wilderness programs and juvenile recidivism: A program evaluation. *Journal of Offender Rehabilitation*, *17,* 19–46 *.

Chiles, A. M. (1984). *Antecedents and outcomes of male delinquents referred to the Utah State Youth Development Center or survival training*. Unpublished doctoral dissertation, Brigham Young University, Salt Lake City, Utah*.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (2nd ed.). Hillsdale, NJ: Erlbaum.

Elrod, H. P., & Minor, K. I. (1992). Second wave evaluation of a multi-faceted intervention for juvenile court probationers. *International Journal of Offender Therapy and Comparative Criminology*, *36,* 247–262 *.

Gass, M. A. (1993). Foundations of adventure therapy. In: M. A. Gass, *Adventure therapy: Therapeutic applications of adventure programming* (pp. 3–10). Dubuque, IA: Kendall Hunt Publishing Co.

Greenwood, P. W., & Turner, S. (1987). *The VisionQuest Program: An evaluation*. Santa Monica, CA: Rand Corporation *.

Hattie, J., Marsh, H. W., Neill, J. T., & Richards, G. E. (1997). Adventure education and Outward Bound: Out-of-class experiences that make a lasting difference. *Review of Educational Research*, *67,* 43–87.

Hawkins, J. D., & Catalano, R. F. (1992). *Communities that care*. San Francisco: Jossey-Bass.

Hawkins, J. D., Herrenkohl, T., Farrington, D. P., Brewer, D., Catalano, R. F., & Harachi, T. W. (1998). A review of predictors of youth violence. In: R. Loeber, & D. P. Farrington, *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 106–146). Thousand Oaks, CA: Sage Publications.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hileman, M. A. (1979). *An evaluation of an environmental stress-chal-*

[1] References marked with an asterisk indicate studies included in the meta-analysis. (Including Kelly and Baer (1968); Langsner and Anderson (1987); Minor and Elrod (1990; 1994); Nunley (1983); Pommier and Witt (1995); Porter (1975); Winterdyk and Roesch (1982); Witman (1987); Zwart (1988) not listed in the text).

lenge program on the social attitudes and recidivism behavior of male delinquent youth. Unpublished master's thesis, Southern Illinois University, Carbondale*.

Huizinga, D., Loeber, R., & Thornberry, T. (1992). *Urban delinquency and substance abuse*. Washington DC: Office of Juvenile Justice and Delinquency Prevention, US Department of Justice.

Kelly, F. J., & Baer, D. J. (1968). *Outward Bound schools as an alternative to institutionalization for adolescent delinquent boys*. Boston, MA: Outward Bound, Inc *.

Kelly, F. J., & Baer, D. J. (1971). Physical challenge as a treatment for delinquency. *Crime and Delinquency*, *17,* 437–445 *.

Kimball, R. O., & Bacon, S. B. (1993). The wilderness challenge model. In: M. A. Gass, *Adventure therapy: Therapeutic applications of adventure programming* (pp. 11–41). Dubuque, IA: Kendall Hunt Publishing Co.

Langsner, S. J., & Anderson, S. C. (1987). Outdoor challenge education and self-esteem and locus of control of children with behavior disorders. *Adapted Physical Activity Quarterly*, *4,* 237–246 *.

Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In: T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller, *Meta-analysis for explanation: A casebook* (pp. 83–127). New York: Russell Sage Foundation.

Lipsey, M. W., & Derzon, J. H. (1998). Predictors of violent or serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In: R. Loeber, & D. P. Farrington, *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 86–105). Thousand Oaks, CA: Sage Publications.

Lipsey, M. W., & Wilson, D. B. (1996). *Toolkit for practical meta-analysis*. Cambridge, MA: Human Services Research Institute.

Minor, K. I. (1988). *An evaluation of an intervention program for juvenile probationers*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo*.

Minor, K. I., & Elrod, H. P. (1990). The effects of a multi-faceted intervention on the offense activities of juvenile probationers. *Journal of Offender Counseling, Services, and Rehabilitation*, *15,* 87–108 *.

Minor, K. I., & Elrod, H. P. (1994). The effects of a probation intervention on juvenile offenders' self-concepts, loci of control, and perceptions of juvenile justice. *Youth and Society*, *25,* 490–511 *.

Molof, M. J. (1967). *Forestry camp study: Comparison of recidivism rates of camp-eligible boys randomly assigned to camp and to institutional programs*. Sacramento, CA: California Youth Authority *.

Nunley, G. L. (1983). The effects of a therapeutic outdoor program on the locus of control and self-concept of troubled youth. Unpublished doctoral dissertation, Oklahoma State University, Stillwater*.

Pommier, J. H., & Witt, P. A. (1995). Evaluation of an Outward Bound school plus family training program for the juvenile status offender. *Therapeutic Recreation Journal*, *29,* 86–103 *.

Porter, W. W. (1975). *The development and evaluation of a therapeutic wilderness program for problem youth*. Unpublished master's thesis, University of Denver, Colorado*.

Ramirez, M. V. (1976). *Evaluation of the San Diego County probation department program — Operation Summit: Final report*. San Diego: San Diego County Probation Department *.

Rohling, M. P. (1987). *An investigation into participant characteristics and goal attainment of the Underway/Spectrum Project*. Unpublished master's thesis, Southern Illinois University, Carbondale*.

Sachs, J. J., & Miller, S. R. (1992). The impact of a wilderness experience on the social interactions and social expectations of behaviorally disordered adolescents. *Behavioral Disorders*, *17,* 89–98 *.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In: H. Cooper, & L. V. Hedges, *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

Weeks, S. Z. (1985). *The effects of Sierra II, an adventure probation program, upon selected behavioral variables of adolescent juvenile delinquents*. Unpublished doctoral dissertation, University of Virginia, Charlottesville*.

Wichmann, T. F. (1990). *Interpersonal problem solving and asocial behavior in a therapeutic wilderness program*. Unpublished doctoral dissertation, Southern Illinois University, Carbondale*.

Wichmann, T. (1993). Of wilderness and circles: Evaluating a therapeutic model for wilderness adventure programs. In: M. A. Gass, *Adventure therapy: Therapeutic applications of adventure programming* (pp. 347–356). Dubuque, IA: Kendall Hunt Publishing Co.

Willman Jr, H. C., & Chun, R. Y. F. (1973). Homeward Bound: An alternative to the institutionalization of adjudicated juvenile offenders. *Federal Probation*, *37,* 52–58 *.

Wilson, J. J., & Howell, J. C. (1993). *A comprehensive strategy for serious, violent and chronic juvenile offenders*. Washington DC: Office of Juvenile Justice and Delinquency Prevention, US Department of Justice.

Winterdyk, J. A. (1980). *A wilderness adventure program as an alternative for juvenile probationers: An evaluation*. Unpublished master's thesis, Simon Fraser University, Toronto, Ontario*.

Winterdyk, J. A., & Roesch, R. (1982). A wilderness experiential program as an alternative for probationers: An evaluation. *Canadian Journal of Criminology*, *24,* 39–49 *.

Witman, J. P. (1987). The efficacy of adventure programming in the development of cooperation and trust with adolescents in treatment. *Therapeutic Recreation Journal*, *21,* 22–29 *.

Ziven, H. S. (1988). *The effects of the Challenge Group treatment program on psychiatrically hospitalized adolescents*. Unpublished doctoral dissertation, Massachusetts School of Professional Psychology, Boston*.

Zwart, T. J. (1988). *The effects of a wilderness/adventure program on the self-concept, locus of control orientation, and interpersonal behavior of delinquent adolescents*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo*.