A meta-analysis of outdoor adventure programming with adolescents

Dana Cason and H.L."Lee" Gillis

Abstract

Adventure practitioners asked to justify their work with adolescent populations have no one study to point to that statistically sums up major findings in the field. Whether it be a school board, treatment facility, or funding agency, one study is needed which can combine statistics from many studies into a format to show overall effectiveness of adventure programming. This study used the statistical technique of meta-analysis to demonstrate that adolescents who attend adventure programming are 62% better off than those who do not. While combining various populations and outcomes resulted in an overall effect that could be considered small by some accounts, the study did point to major problems with current research and offers some direction for future researchers to explore.

In the past 25 years the field of outdoor adventure programming has grown to encompass a wide range of experiential programs relying on challenging physical and mental activities. While some approaches to adventure programming are predominately recreational in nature, others include sophisticated introductions and activity framings geared toward educational or therapeutic goals (Priest & Gass, 1993). Whether the programs are therapeutic, educational, or recreational in nature, adolescents have been a significant part of the participant population. According to Godfrey (1980), adolescents are well suited to adventure programming because of their youthful energy and propensity toward risk-taking. While many practitioners intuitively "know" that adventure programming is effective for

Dana Cason, M.S., works as a marriage and family counselor for Peachbelt Mental Health in Warner Robins, Georgia. Lee Gillis, Ph.D., is an associate professor in the Psychology Department and coordinator of the M.S. track in Adventure Therapy at Georgia College in Milledgeville, Georgia 31061. He also is employed by Project Adventure, Inc., as a licensed psychologist and coordinator of the therapeutic programming team.

adolescents, they are often at a loss when asked for some research data by a board of directors or funding agency to support their claim. Most practitioners find the research that does exist difficult to understand and to have little relevance to their work in the field. This article tries to make numerical sense of the past research in adventure programming with adolescents to offer the practitioner a source for answers to current questions and researchers a direction for future work.

The growth and development of adventure programming has been accompanied by the interest of researchers (who are often practitioners themselves in graduate school) attempting to measure overall program effectiveness. Although generally positive, research results on adventure programming have been contradictory (e.g., Ewert, 1987; Shore, 1977; Wichmann, 1990). The number of research endeavors in adventure programming has grown so large that qualitative literature reviews (like annotated bibliographies) are no longer sufficient tools for understanding the strengths and weaknesses of trying to measure change that results from adventure programming. To date, there has been no attempt to statistically integrate the research findings on adventure programming with adolescents in a way that makes sense to researchers, practitioners, and funding sources. The goals of this investigation were to: (a) identify empirically-based studies concerning adventure programming with the adolescent population, (b) compute and compare overall outcomes of different programs, and (c) relate the outcomes to program characteristics (length, participant population, and study methodology). An examination of the basic relationships among program characteristics and effectiveness will help explain why some studies demonstrated significant improvements while others reported little or no positive outcomes. Findings in this area should promote practitioners to examine the effectiveness of their programs when working with adolescents. Such findings might also help guide new research in adventure programming, especially among masters and doctoral students, towards answering questions that can truly

inform practice and have significant impact on the field. One cannot know the direction for the future, however, without knowing what has happened in the past.

Previous research findings

Adventure programs have been designed around several goals for special populations. Substance abusers, developmentally disabled children, rape and incest victims, sexual perpetrators, psychiatric inpatients, at-risk teens, adjudicated youth, couples, and families represent a few of the populations addressed at annual conferences of the Association for Experiential Education (AEE) or covered in this journal. Naturally, as the implementations of adventure programming have grown, so has the amount of research on its effectiveness.

One area that has been the focus of a considerable amount of study is self-esteem. Some studies revealed no significant differences in pre- and post-measures of self-esteem (e.g., Cytrybaum & Ken, 1975), while others reported significant improvements (e.g., Clifford & Clifford, 1967). Even with inconsistencies, there seemed to be enough evidence to suggest adventure programming led to improved self perception (Shore, 1977). Other researchers have investigated the effects of adventure programming on personality characteristics that predict acting out behavior, locus of control, and depression. Positive findings were more inconsistent with these subjects than they were with self-concept studies.

Research problems confounding the measurement of effectiveness often included a lack of equivalent control groups, a lack of randomization used for participant assignment, a lack of adequate follow-ups on initial successes, and the lack of a clearly defined method for conducting an adventure program (Gillis, 1992). Because of the different foci of adventure programming studies as well as their methodological weaknesses, arriving at an informed opinion regarding effectiveness can be difficult.

In order to address the problem of clarity in the overall analysis of adventure programming effectiveness, a meta-analysis was needed. A meta-analysis is a method of statistically integrating outcomes from many separate studies. The outcomes of each separate study can be converted into an effect size (ES), a measure of the amount of change experienced following an adventure experience. In an adventure programming study involving a control group, the effect size consisted of

the average posttest score of the treatment group minus the posttest score of the control group divided by the standard deviation of the control group (e.g., treatment posttest scores = 150; control group posttest scores = 140 with a standard deviation (SD) of 20; the effect size would be (150 - 140) / 20 = .50). In studies without a control group, the effect size consisted of the average posttest score minus the pre-test score divided by the standard deviation of the posttest score (e.g., 190 on posttest—170 on pre-test divided by SD of 20 = 1.00). Thus, an effect size is the ratio of change due to the program versus change due to chance. An effect size (ES) greater than 0 is positive and denotes an improvement measured in standard deviation units; the higher the ES, the stronger the effect. The .50 from the example above is a 1/2 standard deviation improvement. An effect size less than 0 is negative and denotes a deterioration in the group following a program (or when compared to a control or non-treatment group). An ES of 1.00 is then an improvement of one standard deviation unit (although effects can also be negative if posttest scores are less than pre-test). By computing average effect sizes and standard deviations, comparisons can be made between studies measuring similar constructs (e.g., self esteem) and combinations of studies measuring similar constructs can support general statements about the improvement or deterioration of the variable studied (e.g., adjudicated youth).

A meta-analysis is a method of statistically integrating outcomes from many separate studies. Within the field of adventure programming with adolescents, a meta-analysis was needed to make sense of the large volume of seemingly incongruent research studies.

Meta-analytic methods have been used to evaluate the effectiveness of psychological treatment programs. An analysis conducted by Smith, Glass, and Miller (1980) examined the effectiveness of a over 475 studies comparing treated and untreated groups and revealed an average effect size of 0.85. This means that the average treated person is better off than 80% of the untreated sample. Smith, Glass, and Miller illustrated the clinical meaning of their effect size by contrasting it to effect sizes derived from other studies. For example, in elementary schools, the effects of nine months of instruction in reading is about 0.67 standard deviation units. The increments in mathematics achievement resulting from the use of computer-based instruction is 0.40 standard deviation units. Although there are no set standards against which to evaluate an effect size (ES), Cohen (1977) provided rough guidelines of ES=.2 (small effect size), ES=.5 (medium effect), and ES=.8 (large effect) with the caveat that it is better to obtain comparison standards from the professional literature than to use these somewhat arbitrary guidelines.

With meta-analysis, the effect sizes of one form of

treatment can also be compared with the effect sizes of an alternate treatment. For example, Andrews (1982, 1983) found that psychotherapy involving behavioral treatments with persons diagnosed as having agoraphobia produced a median effect size of 1.30, whereas antidepressant medication with a similar population produced an average effect size of 1.10.

Within the field of adventure programming with adolescents, a meta-analysis was needed to make sense of the large volume of seemingly incongruent research studies. The meta-analysis yielded an average effect size for all adventure programming endeavors. It also offered separate effect sizes for various types of participants, outcome measures, program lengths, and study characteristics so that statistical comparisons could be made. By analyzing effect sizes consistent with the following goals: (a) to identify empirically based studies concerning adventure programming with the adolescent population, (b) to compute and compare overall outcomes of different programs, and (c) to relate the outcomes to program characteristics (length, participant population, and methodology), practitioners, researchers, and funding agencies might be informed as to the relative effectiveness of adventure programming for adolescents.

Method

An attempt was made to collect all available studies of adventure programming with adolescent populations (age 11 through college freshmen) conducted within the last 25 years. Ninety-nine potential studies were identified through literature reviews utilizing Educational Resources Information Center (ERIC), PsychLit, and Dissertation Abstracts International searches. Of the 99 potential studies, 79 were acquired. The 20 that were not acquired were not available through interlibrary loan or references to them were not sufficient for location or acquisition. Of the 79 acquired studies, 36 were excluded from the study because they were: (a) outcome studies that lacked sufficient statistical information for the calculation of effect sizes, (b) not empirically based studies, or (c) did not involve the adolescent population as defined for this research project. Some of the remaining 43 studies reported statistical information on more than one variable (for instance, self-esteem scores as well as behavioral checklist scores); thus the 43 studies used in the analysis generated a total of 235 effect sizes describing 19 outcome measures.

Effect sizes for each outcome of each separate study were computed and averaged to allow for comparison with other studies or for combinations with studies that used a similar variable (e.g., all studies involving adjudicated youth could be compared with each other or all self esteem effects could be combined). Difficulties arose where some studies utilized unique, self designed

outcome measurements. These studies were excluded from the meta-analysis. Some examples of measurements not included were a measurement of dynamic balance and spatial veering for the visually impaired (Black, 1983), selected Jesness Inventory categories (Boudette, 1989), and degrees of recidivism (Chiles, 1984). Other outcome measurements were not used with enough frequency to warrant their inclusion in this study. Since a meta-analysis involves the averaging of effect sizes within a certain category, average effect sizes from outcome measurements used less than five times represented such few studies that they were excluded. In total, 88 effect sizes were excluded from the 235 originally found.

Effect sizes were based on 11,238 adolescents (some adolescents were evaluated on more than one measure). Once the 235 effect sizes were calculated and the 88 effect sizes (based on measurements used less than five times) were disregarded, an average effect size on the remaining 147 effect sizes in the 43 studies for adventure programming with the adolescent population was obtained.

We then compared the effect sizes associated with different program characteristics. The comparison answered questions like: Were the effect sizes different when the adventure programming experience was longer? Were the effect sizes based on studies of delinquent youth different than the ones based on non-delinquent youth? In order to accomplish the comparison, study characteristics had to be defined and coded. The following study characteristics were categorized and coded: (a) duration of the program in hours, (b) participant categories ("normal" adolescents, delinquents, adolescents with emotional or physical handicaps, or population characteristics not specified), (c) average age of the participants, (d) type of outcome measure (e.g., self-concept, locus of control, behavioral measures), (e) date of publication, (f) form of publication (dissertation or journal article), and (g) design rating (a composite score consisting of the sum of the following variables: assignment (random=2, nonrandom=0), timing (pre, post and follow up=2; pre and post=1; post only=0), use of a control group (yes=1, no=0), if follow up was taken, and was it on both treatment and control (yes=1, no=0). The maximum score available was 6).

Once all the data from each study was coded, data analysis consisted of computing average effect sizes, determining correlations among effect sizes and program variables, and employing t-tests to evaluate statistically significant differences. The results are presented and discussed below.

Results and discussion

Average effect size and variance

The adventure outcome effect sizes for the 147 effects in 43 studies ranged from -1.48 to 4.26, with an average effect size (ES) of 0.31 and a standard deviation (SD) of 0.62. The .62 SD represents a large variance in effect sizes reflecting the wide range of study characteristics measured. For example, the effect size of -1.48 came from a dissertation on the changes in self-esteem and locus of control after participation in an Outward Bound program (Stremba, 1977). The effect size of -1.48 means that the 40 participants moved from an internal locus of control (considered by the researcher to be more healthy) toward an external locus of control after the Outward Bound program by a distance of -1.48 standard deviation units. At the other extreme, the effect size of 4.26 came from a study produced for the Australian Outward Bound School in Sydney (Richardson & Richardson, 1982). Twelve inner city males from a Catholic school participated in an Outward Bound program and the effectiveness was measured by their age equivalents on achievement tests. The participants improved by 4.26 standard deviation units on the math component of their tests.

To understand the overall ES of .31 found in this study, consider it as a z-score. By referring to a z-table found in most statistics books, the .31 represents an area under a normal distribution curve of 62.2%. Thus, it could be said the average adolescent participating in an adventure program is better off than 62.2% of adolescents who do not. Another way to understand the effect size involves the amount of improvement shown by the adolescent after the experience. A z-score of 0.31 represents a 12.2% improvement for the average adolescent in the included studies. While some adolescents, especially those in treatment settings improved more than those considered "normal" adolescents the differences were not significantly different; the 12.2% imporvement rate was the average for all the adolescents included in this study.

Compared to a 30% improvement for the average person treated in psychotherapy (Smith, Glass, & Miller, 1980), the 12.2% improvement could be interpreted as meaning adventure programming for adolescents (not adventure therapy per se) is less effective than psychotherapy. However, the studies included in the present meta-analysis differ from the studies included in the meta-analysis by Smith, Glass, and Miller. Combining findings from studies focused on "normal" adolescents with those in a mental health treatment program or in correctional settings may be misleading and one indicator of the wide variation in the overall ES. Such a large standard deviation indicates that studies ranged from very positive effects on some variables to very negative effects on others.

The variation in the evaluation constructs used and measurements taken to gauge program effectiveness was diverse (e.g., self esteem studies were combined with recidivism in one study while a self designed rating was compared with a standardized instrument in another). This variation in findings reflected the lack of standardized, accepted protocols for conducting and measuring the effectiveness of an adventure programming experience. In order to determine a more meaningful estimate of the effectiveness of adventure programming, empirically-based research must develop a more logical and consistent direction. For example, outcome studies reporting little more than the program length, average score of participants, and the resulting change in a variable, whether it be self-esteem or locus of control, are no longer useful in evaluating adventure programming success. More attention must be paid to the such variables as (a) the type of activities utilized, (b) the size of groups (and whether the researcher is studying the group as a unit or studying seperate individuals within a group), (c) the qualifications and characteristics of group leaders, as well as (d) qualitative data and (e) regression data that can help predict who is more likely to be successful in adventure programming (Gillis, 1992).

The present study represents the first research attempt dealing exclusively with adventure programming (only some of which was therapeutic in nature) utilizing an adolescent population. An alternate explanation for the small to medium overall effect size could be that quantitative-based research instruments in the field have not been sensitive to qualitative changes that occur during adventure programming. However, the future appears brighter for such empirically based quantitative assessment since a correlational analysis reveals adventure programming studies conducted in recent years are likely to have higher effect sizes than earlier studies (r=0.18, p=0.004). This finding suggests researchers may be choosing more sensitive evaluation instruments than their earlier colleagues.

Variables affecting effect sizes

It is possible to place undue emphasis on a single summary effect size, especially one with such wide

> Outcome studies reporting little more than the program length, average score of participants, and the resulting change in a variable, whether it be self-esteem or locus of control, are no longer useful in evaluating adventure programming success.

variation, and in light of Cohen's (1977) caveat that it is better to obtain comparison standards from the professional literature than to use arbitrary guidelines of small, medium, and large effect sizes. An asset of this meta-analysis is its ability to examine effect size differences relative to measurement instruments, program length, research design, and participant characteristics. The 147 effect sizes generated from the 43 studies included in the meta-analysis represented seven broad categories of outcome measurements: 1) self-concept, 2) behavioral assessments, 3) attitude surveys, 4) locus of control scales, 5) clinical scales, 6) grades, and 7) school attendance. Average effect sizes from each of the seven categories were significantly different from each

Outcome Measure*	N	ES	SD
Self-concept	61	.339	.700
Behavioral assessment by others	23	.399	.670
Attitude surveys	19	.457	.238
Locus of control	13	.302	.639
Clinical scales	12	1.047	.459
Grades	10	.609	1.527
School attendance	9	.469	.476

N = Number of Effects

* Statistically significant differences found on this construct

Table 1: Average Effect Sizes (ES) and Standard Deviations (SD) for Specific Outcome Measures

other and can be seen in Table 1. The number of effects (N) are listed for each measure as well as the average effect size (ES) and the standard deviation (SD) for each outcome measure. Clinical scales (e.g., MMPI), grades, school attendance, and attitude surveys had the highest effect sizes. The effect sizes for specific outcome measurements reveal that adventure programming affected scores on clinical scales (e.g., depression and anxiety) more than locus of control measures (1.047 compared with .302).

A possible explanation for the elevated average effect size in the category of clinical scales is that the majority of these effect sizes emerge from studies with residential populations in treatment centers or adjudicated youth in alternative correctional placements. Residential participants are perhaps more likely to experience adventure programming as a part of an adjunctive or primary therapy process conducted by mental health professionals than as a purely educational or recreational experience (Gillis, Gass, Bandoroff,

Rudolph, Clapp, & Nadler, 1991) . Accordingly, residential participants' adventure experience could be more emotionally intense than the adventure experience of someone participating in a recreational or educational program. Alternatively, such participants could scorer lower than "normals" at pre-test and then "improve" to score higher (but still within a "normal" range at posttest) resulting in a larger overall effect size for treatment populations than for "normals."

An important variable examined in this meta-analysis was program duration. A significant positive correlation was found between the length of the program and the effect size ($r=0.174$, $p=0.008$). Program lengths were measured in hours, and groups spending consecutive days in a wilderness setting were estimated to have spent 18 hours per day in adventure programming. Durations ranged from 36 to 5400 hours (ten months) with a median length of 54 hours (three weeks). Three weeks was considered a moderate length and was the program duration for 41% of the outcome measurements taken (mostly from Outward Bound or Outward Bound-type programs). Shorter programs represented 27%, and longer programs represented 32% of the sample. Statistical analysis revealed significant differences between effect sizes associated with longer programs when compared with effect sizes associated with shorter or moderate length programs. A survey of substance abuse treatment programs found the mode of treatment for that population to be one day (Gass & McPhee, 1990). The results of this meta-analysis would suggest that adventure programs are more effective if they are longer, however, this analysis was unable to determine an optimal length of adventure programming.

The age of the adolescents participating in the study was negatively linked with effect size suggesting younger participants benefited slightly more than older participants from adventure programming ($r=-0.18$, $p=0.01$). The average age of the adolescents who participated was 15.8 with a standard deviation of 0.92.

Participants in the meta-analysis belonged to different population groups: adjudicated youth, inpatients, emotionally or physically challenged, at-risk adolescents (defined by school officials) and "normal" adolescents. Statistical analysis revealed no significant differences in effect sizes resulting from the various groups; "normal" adolescents were just as successful as diagnosed or other "labled" populations in this analysis of adventure programming. Related to the earlier finding of differences between specific outcome measures, "clinical" or treatment populations often were given a battery of evaluation instruments while "normal" populations were usually evaluated on fewer instruments (e.g., personality inventories, self esteem instruments, and recidivism for treated populations while "normal" populations only evaluated on self esteem)

A final variable examined in the meta-analysis was the research design. Each study was rated on a scale

om 1 to 6 with "1" being the most informal, requiring
nly pre- and posttesting without a control group; "6"
:presented the most scientifically sound studies
ivolving a control group with random assignment and
re-, post and follow-up testing of both experimental
nd control groups. Most effect sizes in the meta-analy-
is (70.6%) emerged from studies <u>without</u> randomly
ssigned control groups. The research design rating cor-
related negatively with effect sizes, revealing that stud-
:s relying on more rigorous measures of effectiveness
ad lower effect sizes than studies with loosely defined
leasure techniques ($\underline{r}$=-0.28, $\underline{p}$=0.01). This finding
vould indicate that studies not as empirically sound
vere more likely to have more positive findings,
ncreasing speculation that they are perhaps attributing
hange in their findings when such change is due to
hance (a Type I error in statistics).

In order to further investigate the relationship
between research design and effect sizes, studies were
divided into four classes and are shown in Table 2.
Studies with ratings of "1" were categorized as "poor";
studies with ratings of "2" and "3" were categorized as
"fair"; studies with ratings of "4" were categorized as
"good"; and, studies with ratings of "5" and "6" were
categorized as "excellent." Table 2 reveals average
effect sizes resulting from various combinations of
design quality, participant categories, and program
duration. Studies categorized as poor produced signifi-
cantly higher effect sizes than studies categorized as
fair, good, or excellent. Studies categorized as fair pro-
duced significantly higher effect sizes than studies cate-
gorizes as excellent. The more stringent the study, the
lower the effect size. In addition, different populations
produced a variety of results including negative find-
ings for emotionally and physically chal-
lenged students in studies with excellent
designs to very positive findings for the
same population in poorly designed stud-
ies. In all of the findings in Table 2, keep
in mind that the number of effects in each
category and the wide variation of find-
ings (indicated by the higher standard
deviations) resulted in comparisons
between populations that were statistical-
ly insignificant; only length of program
and type of design produced statistically
significant findings.

Might the results in Table 2 imply
that the effects of adventure programming
are negligible when well-designed studies
are utilized? Might the well-designed
studies involve measurement instruments
that are not sensitive to the changes that
might be occurring during adventure pro-
gramming? Might the level of training of
the leaders involved in conducting these
various programs be so different that it
accounts for the lack of significant find-
ings among populations by contributing
to much of the measurement error in the
findings? While these questions can be
raised, they cannot be answered by the
present study. Such seemingly important
variables like qualifications and qualities
of leaders were not addressed in studies
with enough consistency to make coding
of this variable possible. The only other
criteria that could be assessed related to
whether the study was published in a ref-
ereed journal or was part of a dissertation
or non-refereed publication.

Design*	Population	Length*	N	ES	SD
Excellent	E/P prob	Short	8	-.283	.267
	JD	Moderate	48	.132	.351
	JD	Short	4	.632	1.038
	NO	Moderate	9	.110	.192
Good	JD	Long	20	.254	.679
	NA	Moderate	4	-.002	.340
	NO	Short	9	.127	.399
Fair	E/P prob	Long	4	.743	.208
	JD	Long	4	.282	.087
	JD	Moderate	32	.262	.307
	NA	Long	4	.486	.202
	NA	Short	4	-.198	.436
	NO	Long	6	.147	.916
	NO	Short	11	.670	.613
Poor	E/P prob	Long	6	1.109	.382
	E/P prob	Moderate	3	.497	.219
	JD	Long	26	.738	.567
	JD	Moderate	1	1.436	-
	JD	Short	13	.031	.124
	NA	Long	5	1.092	2.26
	NO	Short	4	.148	.153

E/P emotionally or physically challenged
JD adjudicated youth
NA population description was unclear or not specified
NO normal adolescents
N = Number of Effects
* Statistically significant differences found on this construct

Table 2: Average Effect Sizes (ES) and Standard Deviations
(SD) for Level Combinations of Design, Sample,
Population, and Time Variables

The average effect sizes for published and unpub-
lished studies are presented in Table 3. Effect sizes
were significantly higher in refereed journal articles

than in dissertations (t=2.29, p=0.03). This finding is not surprising since many journals typically do not publish studies that demonstrate no differences between groups or have "negative" findings. One might make the inference that more poorly designed studies are being published more often since both poorly designed and published studies each have higher effect sizes. What is probably more likely is that the unpublished studies represent some of the best and worst of the empirical research in adventure programming. Indeed, a glance at Table 3 demonstrates that most of the research utilized in this study is unpublished. The table points to one of the most glaring problems in adventure programming research: the lack of published studies compared to those conducted.

Design	Unpublished			Published		
	N	ES	SD	N	ES	SD
Excellent	65	.077	.348	4	.632	1.038
Good	43	.215	.627	0	-	-
Fair	60	.345	.471	5	.243	.727
Poor	42	.517	.777	16	.844	.916

N = Number of Effects

Table 3: Average Effect Sizes (ES) and Standard Deviations (SD) for Published and Unpublished Studies Grouped by Design Rating

Summary and conclusions

In summary, the average of the 147 effect sizes collected for the meta-analysis of adventure programming with adolescents was 0.31. This finding represents a 12.2% improvement for the average adolescent participating in an adventure program. Adolescents who participate in adventure programming are better off than 62% who do not participate. According to Cohen (1977), 0.31 is a small to moderate effect size. However, when one considers the large range of effect sizes collected, -1.48 to 4.26, and the wide variation in the findings, a singular summary effect size focused exclusively on quantitative analysis of primarily dissertation-based studies cannot adequately reflect the effectiveness of adventure programming.

Summary effect sizes of outcome measurement categories (e.g., self concept, locus of control, clinical scales) were significantly different from each other and ranged from 0.30 to 1.05. Longer programs and younger participants were linked with larger effect sizes. As study designs approached "true" experiments, effect sizes decreased, and published studies produced significantly higher effect sizes than unpublished dissertations.

Adventure programming was not shown to be significantly more effective with adjudicated adolescents than it was with other adolescent populations; it was equally effective.

The limitations of the present meta-analysis reflect the limitations of existing research in the relatively new field of evaluating adventure programming. Many potentially important variables are not routinely documented in the research. Leadership training and leadership styles are rarely taken into account in a description of a study and thus it is difficult to know whether poorer results were found from leaders who were less well trained or had a leadership style that did not match the group. Likewise, few details were given about characteristics of the participants in the studies beyond traditional demographic information. Researchers trying to understand which type of participant will do best in which type of program need to specify more information about those who participate. Finally, specific activities utilized, the type of facilitation style employed, and the order in which activites were presented, plus any time spent processing experiences, were rarely discussed. Often the activities chosen and how they are processed are what practitioners want most from research studies. Researchers must do a better job of accurately describing what occurs during the adventure programming sessions in order to make research more valid. Such detailed descriptions of effective programs can allow for further replication of positive findings.

The present meta-analysis is an attempt to statistically integrate all available empirical research on adventure programming with adolescents. The studies varied a great deal in their designs, methods of presentation, and even in their goals. Adventure programming experiences ranged from college courses in outdoor activities to three-week Outward Bound experiences. The inclusion of a wide variety of adventure programs may have obscured some important questions concerning the most effective adventure experience with the most effected populations. However, the study does represent the first attempt to statistically evaluate a number of studies conducted with a similar age group. The wide variance in findings raises questions about the validity of quantitative research for this field, the reliability of instruments used for assessment of pre- and postprogram changes, and the host of unknown variables that may be influencing both positive and negative effects of adventure programming. The more we can learn from each others' successes and mistakes through program evaluation and research, the more we can benefit our clients and our selves in using adventure programming.

Note

The authors would like to thank the following people for helpful pre-publication advice and counsel: Dene and Jenny Berman, Michael Gass, Simon Priest, John Lindsay, and Brian Martin, as well as the review process of the *Journal*.

References

Andrews, G. (1982). A methodology for preparing "ideal" treatment outlines in psychiatry. *The Australian and New Zealand Journal of Psychiatry, 16,* 153-158.

Andrews, G. (1983). A treatment outline for depressive disorders. *Australian and New Zealand Journal of Psychiatry, 17,* 129-146.

Black, B. C. (1983). The effect of an outdoor experiential adventure program on development of dynamic balance and spatial veering for the visually impaired. *Therapeutic Recreation Journal, 17*(3), 39-49.

Boudette R. D. (1989). The therapeutic effects of Outward Bound with juvenile offenders. *Dissertation Abstracts International, 50*/11-B, 5306. (University Microfilms No. AAD89-26357132)

Chiles, A. M., Jr. (1984). *Antecedents and outcomes of male delinquents referred to the Utah State Youth Development or survival training.* Doctoral Dissertation, Brigham Young University, 1984.

Clifford, E., & Clifford, M. (1967). Self-concepts before and after survival training. *British Journal of Sociology and Clinical Psychology, 6,* 241-248.

Cohen, J. (1977). *Statistical power analysis for behavioral sciences* (revised ed.). New York: Academic Press.

Cytrynbaum, S., & Ken, K. (1975). *The Connecticut Wilderness Program: A preliminary evaluation report.* Hartford Connecticut Council on Human Services.

Ewert, A. (1987). Research in outdoor adventure: Overview and analysis. *Bradford Papers Annual, 2,* 15-28.

Gass, M. A., & McPhee, P. J. (1990). Emerging for recovery: A descriptive analysis of adventure therapy for substance abusers. *Journal of Experiential Education, 13*(2), 29-35.

Gillis, H. L. (1992, January). Therapeutic uses of adventure-challenge-outdoor-wilderness: Theory and research. In K.A. Henderson (Ed.), *Coalition for Education in the Outdoors Research Symposium Proceedings.* Bradford Woods, IN: ERIC Document Reproduction Service, ED 352227.

Gillis, H. L., Gass, M. A., Bandoroff, S. Rudolph, S. Clapp, C., and Nadler, R. (1991). Family adventure survey: Results and discussion. In C. Birmingham (Ed.) *Proceedings Journal of the 19th Annual AEE Conference.* Boulder, CO: Association for Experiential Education, pp. 29-39. (ERIC Document Reproduction Service, ED 342584)

Godfrey, R. (1980). *Outward Bound: Schools of the possible.* Garden City, NY: Anchor Press/Doubleday.

Priest, S. & Gass, M. A. (1993). Five generations of facilitated learning through adventure. *Journal of Adventure Education and Outdoor Leadership 10(3),* 23-25.

Richards, G. E., & Richards, M. F. J. (1981). *Outward Bound Bridging Course 1981: An investigation and evaluation of an Outward Bound Remedial Programme.* Sydney: Australian Outward Bound School. (ERIC Document Reproduction Service No. ED 225 799)

Shore, A. (1977). *Outward Bound: A reference volume.* Greenwich, CT: Outward Bound, Inc., Greenwich, CT: Outward Bound, Inc.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy.* Baltimore: Johns Hopkins University Press.

Stremba, R. H. (1977). A study of the relationship between participation in an Outward Bound program and changes in self-esteem and locus-of-control. *Dissertation Abstracts International.* (University Microfilms No. 77-27014)

Wichmann, T. F. (1990). Interpersonal problem-solving and asocial behavior in a therapeutic wilderness program (adventure programs). *Dissertation Abstracts International, 52/05-A.* (University Microfilms No. AAD91-29891).