DOCUMENT RESUME

ED 251 271 RC 015 060

AUTHOR Marsh, Herbert W.; And Others
TITLE Multidimensio · Self-Concepts: The Effect of Participation an Outward Bound Program.
PUB DATE 29 Sep 84
NOTE 29p.; Request for reprints should be sent to Dr. Herbert W. Marsh Dept. of Education, University of Sydney, Sydney, NSW 2006, Australia.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adolescents; Adults; *Adventure Education; Foreign Countries; Locus of Control; *Outcomes of Education; *Outdoor Education; *Participant Satisfaction; *Program Effectiveness; *Self Concept; Self Concept Measures
IDENTIFIERS Australia; *Outward Bound

ABSTRACT

Questionnaires were administered to 361 participants of a standard 26-day Australian Outward Bound Program to examine systematic change and stabiltiy in multiple dimensions of self-concept, to test hypothesized effects of Outward Bound participation on self-concept, and to explore methodological issues in such studies. The participants (aged 16-31, 75% male) had taken 1 of 10 standard residential Outward Bound courses offered between November 1982 and May 1983. Self-description questionnaires were completed 1 month before the start of the program, on the first day of the program, and on the last day of the program. Findings showed that participation in the program produced increases in the multiple dimensions of self-concept over the 26-day interval, demonstrating the program's effectiveness. Counter explanations for the findings were examined with a variety of different approaches and did not appear to be viable. The psychometric properties of responses to the self description questionnaire (reliability, dimensionality, and stability), coupled with the systematic relationship between the size of shifts in the questionnaire scales and the scales' a priori relevance to program goals, supported the validity of interpretations based upon the questionnaire and its use as a criterion measure in intervention studies. (Author/NEC)

Multidimensional Self-concepts: The Effect of Participation in an Outward Bound Program

Herbert W. Marsh
The University of Sydney

Garry E. Richards, Director
Australian Outward Bound School

Jennifer Barnes
The University of Sydney

7 June, 1984
Revised: 29 September, 1984

Running Head: Outward Bound

Acknowledgments

The authors would like to acknowledge Samuel Ball, Raymond Debus, and Lee Owens for their helpful suggestions at various stages of this research.

Requests for reprints should be sent to Dr. Herbert W. Marsh, Department of Education, The University of Sydney, Sydney, NSW 2006, Australia.

Multidimensional Self-concepts: The Effect of Participation in an Outward Bound Program

ABSTRACT

The purposes of this study are to examine systematic change and stability in multiple dimensions of self-concept, to test hypothesized effects of participation in the Outward Bound Program on self-concept, and to explore methodological issues in such studies. A total of 27 groups (N= 361 participants aged 16 to 31, 75% male) participated in a 26-day residential program. Participants completed the Self Description Questionnaire III (SDQ III) one month before the start of the program (time 1), on the first day of the program (time 2), and on the last day of the program (time 3). Participation in the program produced increases in the multiple dimensions of self-concept over the 26-day interval, demonstrating the program's effectiveness. Counter explanations for the findings were examined with a variety of different approaches, and did not appear to be viable. The psychometric properties of responses to the SDQ III -- reliability, dimensionality, and stability -- coupled with the systematic relationship between the size of shifts in SDQ III scales and the scales' a priori relevance to program goals, support the validity of interpretations based upon the SDQ III and its use as a criterion measure in intervention studies.

Multidimensional Self-concepts: The Effect of Participation in the Outward Bound Program

Self-Concept.

Enhancement of self-concept is widely valued as a desirable goal and has been posited as an intervening process that may lead to changes in other outcomes. However, systematic reviews of self-concept research emphasize the lack of theoretical basis in most studies, the poor quality of measurement instruments used to assess self-concept, methodological shortcomings, and a general lack of consistent findings (e.g., Burns, 1979; Shavelson, Hubner & Stanton, 1976; Shavelson & Marsh, in press; Wells & Marwell, 1976; Wylie, 1974; 1979). Self-concept, like many other psychological constructs, suffers in that "everybody knows what it is" and researchers do not feel compelled to provide any theoretical definitions of what they are measuring. In an attempt to remedy this situation, Shavelson et al. (1976) posited a multifaceted, hierarchical model of self-concept. While the multidimensionality of self concept is not universally accepted, the assumption has received strong support in a recent review of research stimulated by the Shavelson model (Shavelson & Marsh, in press), and from research by other investigators (e.g., Boersma & Chapman, 1979; Dusek & Flaherty, 1981; Fleming & Courtney, 1984; Harter, 1982; Soares & Soares, 1982). In their review Shavelson and Marsh contended that the relationship between self concept and other constructs cannot be adequately understood if this multidimensionality is ignored.

Construct validity studies in self-concept can be classified as within-network or between-network research. Within-network studies test the existence of hypothesized facets of self-concept (e.g., physical, social and academic) through procedures like factor analysis and multitrait multimethod analyses. Between-network studies attempt to demonstrate a theoretically consistent pattern of relations between self concept and external criteria such as other self report measures, ability/achievement indicators, self-concept ratings inferred by significant others, behavior, and the impact of systematic interventions that are designed to enhance self-concept. Increasingly, researchers incorporate both these emphases, attempting to demonstrate that external criteria are systematically related to particular facets of self concept to which they are most logically related, and substantially less related to other facets of self-concept. For example, in an analysis of the self

concept/academic achievement relationship. Marsh (1984) reported that verbal and mathematical achievements were substantially correlated with Verbal and Math self-concepts, but were nearly uncorrelated with nonacademic areas of self-concept.

Stability and Change in Self-Concept. Self-concept investigators face an interesting dilemma. Theoretically, and from the perspective of measurement theory, it is desirable that self-concept be relatively stable over time. Since general self-concept, and even particular facets of self-concept, are based on such a wide range of experience, even dramatic life events and interventions may have only a modest effect on self-concept. Some researchers also argue that it is important to a person's mental health that self-concept be relatively stable. Nevertheless, much of the interest in self concept is directed toward changes in self-concept, as is the case in most intervention studies that employ self-concept measures. Self-concept cannot be perfectly stable and still be responsive to dramatic life events or systematic interventions, and herein lies the dilemma.

Marsh, Smith, Barnes and Butler (1983) examined changes in multidimensional self-concepts of preadolescent children in a study that has particular relevance to this dilemma. First, their research identified multiple facets of self concept that were derived from the Shavelson model and showed that these multiple facets were systematically related to self concepts inferred by teachers and to academic achievement indicators. Second, they found that the facets of self concept were relatively stable, even for young children. Third, they found that changes in self concept that did occur were reliable and systematic. Finally, their results suggested that changes in self concept were multidimensional such that changes were specific to particular dimensions of self concept rather than generalizing across all dimensions. Thus, they suggested, it is possible for an intervention to have a moderate effect on some particular facet of self concept, even if it has a less substantial effect on overall self concept. These results may offer an acceptable compromise to both sides of the dilemma; self concept is relatively stable, but the changes that do take place are reliable and specific to particular dimensions of self concept. However, these researchers were not able to demonstrate that the systematic and reliable changes in multiple facets of self concept were substantially correlated with changes in other constructs, and they did not examine changes in self concept produced by a systematic intervention.

The study of changes in multiple dimensions of self-concept due to systematic interventions has important practical and theoretical implications, but is particularly plagued with methodologically weak studies, with studies based on poor measurement instruments, and with a paucity of consistent findings. While a systematic review of such research is beyond the scope of this article, reviews by Wylie (1979) and by Scheirer and Kraut (1979) are particularly relevant. Wylie reviewed research proposing to alter self-concept as a consequence of psychotherapy. She concluded that an "overview of all the substantively summarized research gives no support for the belief that allegedly therapeutic or growth-producing group experiences affect the over-all level of self-regard of volunteer 'normal,' young adult participants" (1979, p.642-643). Scheirer and Kraut (1979) reviewed the effects of academic interventions on self-concept and academic achievement; they also failed to find consistent pattern of results or many studies that produced any change in self-concept. Wylie cautioned that there are many reasons why we should be suspicious of systematic changes in self-concept produced by interventions (e.g., placebo effects, acquiescence to the experimenter, etc.). However, when the trends are not positive despite these apparent biases, then the hypothesized changes in self-concept are particularly tenuous. With regard to psychotherapy research she suggested two alternatives; to reject the popular and plausible notion that therapy should improve self-concept, or to assert that the notion is too intuitively plausible to abandon in the face of null findings and to pursue methodologically and theoretically more sophisticated research.

Why doesn't intervention research demonstrate more systematic changes in self-concept? Many probable reasons exist, but two are particularly relevant to the present investigation. First, most research uses ill-defined measures of overall self-concept that ignore the multidimensionality of the construct. If self-concept is multidimensional, as asserted in this study, then some areas of self-concept will be logically related to the goals of an intervention while others will not. Second, the size of the likely effect relative to the probable error is typically small. This will be the case when the intervention is weak or when the number of subjects in the study is small. The primary purpose of the present investigation is to examine the effect of a powerful intervention, the Outward Bound program described below, on multiple dimensions of

self-concept for a large sample of participants.

The Outward Bound Program.

The first Outward Bound School was developed by Kurt Hahn in Wales in 1941 as a way of training young British seamen to withstand the hazards of naval warfare; the initial program was designed to allow young men to realise their potential, and to develop a stronger character and a will to survive. Since 1941 more than 35 Outward Bound Schools have been established in more than 15 countries. The standard Outward Bound course is a 26-day residential program for young men and women 17 to 29 years of age (or older with special permission). The course is both physically and mentally very demanding. Participants enroll in the course for many reasons including the enjoyment and challenge of the outdoors activities, broadening their outlook on life, developing character, and enhancing leadership potential. Some are sponsored by employers; some are sent by their parents; some are sponsored by social service organizations; and some sponsor themselves. Historically, Outward Bound participants have been young men, but an increasing number of young women are participating in mixed-sex or all-female groups.

Participants work in groups of about 12, sharing most of the activities, many of which require considerable group organization and cooperation among the individuals. As part of the program each person follows a personal physical fitness program which includes cross country runs, rock climbing, and other vigorous activities. Participants travel up to 300 km on foot in rough mountainous terrain, or paddling their own craft in swiftly flowing rivers. While the strong emphasis on physical, outdoor activities is a part of Outward Bound, physical fitness is not the primary goal. Instead, Richards argues that the physical activities are an effective medium for "the person to recognize and understand his own weaknesses, strengths, and resources and thus find within himself the wherewithal to master the difficult and unfamiliar" (1977, p.69). Similarly, overt, external competition is deemphasized and the focus is on "the internal competition of the individual with himself, trying to achieve greater personal goals, ... which form the basis of competition between where the individual's standards have been, and where he wants them to be" (Richards, 1977, p. 93). The establishment and fulfillment of personal and group goals in outdoor physical activities, the group experience, and the opportunity to experience and master stressful situations are

important components of the program. In addition to increased physical fitness, goals of the program include improved awareness of self, self-confidence, initiative and self-reliance, cooperation, awareness of others, taking responsibility for self, and the mastery of stressful situations (see Richards, 1977, for further discussion of the standard course and the aims of the Outward Bound course examined in this study).

Outward Bound Research Studies. A large amount of research from all over the world has been conducted to evaluate the impact of Outward Bound programs and these have been reviewed by Godfrey (1974), Richards (1977), and Shore (1977). Shore (1977) noted that most of the research is unpublished and methodologically flawed, leading him to argue that "One must conclude, overall, that the research literature of Outward Bound is weak." None of the 80 plus studies reviewed by Shore used a true experimental design, and few systematically examined threats to the validity of interpretations based on pretest-posttest difference scores. In one of the most carefully considered studies, Smith, Gabriel, Schott and Padia (1975, p.4) recognized the potential value of a randomly assigned control group but argued that: "Interference by the researchers into participant selection and placement necessary in implementing a true experimental design was impossible for this study, however. The decision was made instead to use the interrupted time-series design" (p.4). On the basis of their study they found evidence that the program had a positive impact on self assertion (taking personal responsibility and actively confronting new or stressful situations) and on self esteem.

Self concept is the most frequently examined psychological construct in the Outward Bound literature. Richards (1977) reviewed theoretical and empirical bases for the impact of Outward Bound on self-concept and concluded that "The observations of so many tens of thousands of students throughout the world who have completed the Outward Bound courses clearly support the influence Outward Bound has on the individual's concept of self. These observations have also gained increasing support, especially since the mid 1960's from a growing body of research" (p.96). While many of the studies in Shore's 1977 review also reported that participation in Outward Bound leads to a more positive self-concept, he concluded that the findings were mixed, methodologically flawed, and often confused. However, important problems in the measurement of self concept

discussed earlier mean that it is difficult to disentangle measurement problems from the impact of the program. Hence it is not surprising that Shore found some methodological and empirical inconsistencies when reviewing self-concept studies in the Outward Bound literature. An important focus of the present study is to examine psychometric properties of a well developed self-concept instrument and the impact of the Outward Bound program on responses to this instrument.

METHOD

Measurement Instruments.

In addition to the background/demographic information normally collected by the Outward Bound Program, all participants were required to complete the Self Description Questionnaire (SDQ) III and the Rotter Locus of Control (LOC) scale described below.

SDQ III. The SDQ III is one of a set of self-report instruments designed to measure self-concepts of preadolescents (SDQ), early-adolescents (SDQ II) and late-adolescents and young adults (SDQ III). The set of SDQ instruments is based upon the Shavelson model of self-concept (Shavelson, Hubner & Stanton, 1976), the facets of self-concept proposed in that model, and earlier research with the SDQ instruments (see Marsh, Barnes & Hocevar, in press; Marsh & Shavelson, 1984). Numerous exploratory and confirmatory factor analyses of responses to the SDQ instruments have identified the factors that each is designed to measure and support the multidimensionality of self concept (e.g., Marsh, Cairns, Barnes & Tidman, 1984; Marsh & O'Neill, 1984; Marsh, Parker & Barnes, in press; Marsh, Relich & Smith; 1983; Marsh, Smith & Barnes, in press; Shavelson & Marsh, in press). Other research with the SDQ instruments has shown that: a) the reliability of each factor is generally in the 0.80's and 0.90's while correlations among the factors are modest (median r's are generally 0.20 or less); b) the self-concept factors are substantially correlated with self concepts in matching areas as inferred by teachers and by significant others (Marsh, Barnes & Hocevar, in press; Marsh & O'Neill, 1984; Marsh, Smith & Barnes, 1983; in press; Marsh, Smith, Barnes & Butler, 1983); and, c) the self-concepts in academic areas are substantially correlated with academic achievement indicators while nonacademic self-concepts are not (Marsh, 1984; Marsh & Parker, 1984; Marsh, Parker & Barnes, in press; Marsh & O'Neill, 1984; Marsh, Parker & Smith, 1983; Marsh, Smith, Barnes & Butler,

1983). These findings support the validity of interpretations based on the SDQ instruments.

The SDQ III, the rationale for its construction, its relation to the Shavelson model and the other SDQ instruments, the wording of the items, its psychometric properties, its relation to academic achievement and to self-concept inferred by significant others, is summarized elsewhere (Marsh, Barnes & Hocevar, in press; Marsh & O'Neill, 1984). Each of the 13 SDQ III scales is represented by 10 or 12 items, approximately half of which are negatively worded, and subjects respond on an eight-point response scale where categories vary from "1--Definitely False" to "8-Definitely true." The 13 scales are described below.

1. Math -- I have good mathematical skills/reasoning ability.
2. Verbal -- I have good verbal skills/reasoning ability.
3. General Academic -- I am a good student in most school subjects.
4. Problem Solving -- I am good at problem solving/creative thinking.
5. Physical Ability -- I am good at sports and physical activities.
6. Appearance -- I am physically attractive/good looking.
7. Relations With Same Sex -- I have good interactions/relationships with members of the same sex.
8. Relations With the Opposite Sex -- I have good interactions/relationships with members of the opposite sex.
9. Relations With Parents -- I have good interactions/relationships with my parents.
10. Religion/spirituality -- I am a religious/spiritual person.
11. Honesty -- I am an honest, reliable, trustworthy person.
12. Emotional Stability -- I am an emotionally stable person.
13. General-Self -- I have self-respect, self-confidence, self-acceptance, positive self-feelings and a good self-concept.

The Rotter LOC scale. Rotter (1966; Rotter, Chance & Phares, 1972; also see MacDonald, 1973) describes the Rotter LOC, its theoretical basis, and empirical studies based upon it. Internal-external locus of control is defined as the extent to which a person perceives reinforcement as contingent upon his/her own behavior or independent of it. The locus is internal when the person perceives that events are contingent upon his/her own behavior and relatively permanent personal characteristics, and external when they are seen as contingent upon luck, fate, the control of powerful others, the environment, or some other characteristic not under his/her own

control.

The Rotter LOC consists of 23 question pairs, using a forced-choice format, and six filler questions. Each pair of questions contains one internal statement and one external statement; the total score is the number of external responses. Rotter (1966; Rotter, Chance & Phares, 1972) presents internal consistency estimates of reliability of about 0.70 from a number of different studies. Test-retest estimates of reliability are nearly as high, though they are more varied and generally based upon much smaller sample sizes. Factor analyses described by Rotter and by others suggest that there is one general factor which accounts for most of the variance in the total score, though MacDonald (1973) suggests that the scale is not as factorially pure as it was once believed to be. Validity research with the Rotter LOC is summarized by Rotter (1966; Rotter, Chance & Phares, 1972) and by Lefcourt (1976).

Subjects, Design and Procedures.

Subjects are the 361 participants, aged 16 to 31 (median age = 21), who completed one of 10 standard Outward Bound courses offered between November 1982 and May 1983 at one of two different Australian sites. 96% of the participants were single and 75% were male. Participants represented a wide range of socioeconomic backgrounds. About one-third of the participants reported that they had been full-time students during the year prior to the course, and 23% said that they would be students the following year. A majority (60%) had been employed on a full-time basis during the year prior to the course. When presented with a list of options of what they hoped to get from Outward Bound, participants indicated: broaden outlook and experience (82%), learn new outdoor skills (59%), develop character/personality (59%), chance to get fit (48%), have an enjoyable time (37%), develop ability to mix socially (27%), and a holiday (15%).

Enrollments for the 10 different courses varied between 13 and 57, and within each course participants were divided into groups of about 12. Of the 27 groups of participants, 15 comprised all males, 3 all females, and 9 were mixed-sex groups in which between 31% and 50% of the members were female. Each course was actually conducted at one of two different Australian sites.

Participants completed the SDO III approximately one month before the start of the course, together with course registration materials (time 1), again on the first day of the course (time 2),

and also on the last day of the 26-day course (time 3). The LOC was completed at times 2 and 3, and background/demographic information normally collected as part of the Outward Bound program was collected at time 2. At time 1 materials were mailed to participants and completed in their homes, while at times 2 and 3 materials were completed at the Outward Bound site. Since the data collection was conducted as part of the program, there were very few missing responses; 348 participants completed the materials all three times and none failed to complete the self-report surveys for more than one of the testing points.

Based upon previous research and the aims of the program described by Richards (1977), it was predicted that participation in Outward Bound would result in more positive self-concepts and a more internal LOC. However, the SDQ III clearly differentiates among 13 facets of self-concept which appeared to have varying amounts of relevance to the program. Consequently, before the start of the data collection, the Program Director was asked to judge the relevance of each SDQ III scale to the aims of the program, and whether the relevance would vary for mixed-sex and single-sex groups. These judgments were incorporated in subsequent analyses.

Statistical Analysis.

Psychometric Properties of the SDQ III and the LOC. In analyses of responses to the SDQ III, responses to negative items were reversed so that a response of "8" always represented the most positive response, indicating a higher self-concept. The median response was then substituted for the few missing values that existed (less than 1/4 of 1%) in completed questionnaires. The commercially available SPSS program (Hull & Nie, 1981) was used to determine the internal consistency estimates for each SDQ III scale separately at time 1, time 2, and time 3, and also for difference scores representing the time 2 to time 3 change in responses. Then the 10 (or 12) items from each scale were divided into 5 (or 6) item pairs such that the first two items in the scale were assigned to the first pair, the next two items to the second pair, and so forth. Factor analyses were performed on the sum of responses to the two items in each item pair as is typically done in other SDQ research (see Marsh & O'Neill, 1984). The factor analyses, using iterated communality estimates and an oblique rotation, following a Kaiser normalization, were performed with the commercially available SPSS program (Nie, et al., 1975). Separate factor analyses were

performed on responses from time 1 (n=357), time 2 (n=358), and time 3 (n=355), and then a combined analysis was performed on all 1,070 sets of responses, with responses by each subject at each time being considered to be a separate case. Factor scores (Nie, et al., 1975) were derived from this combined factor analysis to represent the 13 SDQ III scales at each time separately so that all the scores were based upon a common metric and so that mean differences between different times were not lost. Across all responses from time 1, time 2 and time 3, each of the 13 SDQ III factors was standardized to have mean = 50 and SD = 10. A similar factor analysis was also performed on the difference scores representing the change in time 2 to time 3 responses to the item pairs. Finally stability coefficients were computed for the time 1 to time 2, time 1 to time 3, and time 2 to time 3 intervals.

Responses to the dichotomously scored LOC items were scored so that "1" reflected an external response and "0" an internal response; the total score represented the number of external responses. Internal consistency estimates were determined separately for LOC responses for time 2 and for time 3, and the stability coefficient was determined for the time 2 to time 3 interval.

Program Impact. The nature of the Outward Bound program, a 26-day residential program scheduled at specific times of the year in different parts of the country, meant that randomly assigned No Treatment and Placebo control groups were not feasible. Thus, the design of the study is a critical feature. For purposes of this study, the design is characterized as a multiple interrupted time series. Textbooks dealing with the statistical analysis of time series typically suggest testing at least 50 points in time, instead of just three. However, as emphasized by Cook and Campbell (1979), the advantages of using even a "short" time series are important, even if Box-Jenkins types of analyses are not appropriate. We choose to use the term time-series to emphasize the difference between the present design and the typical one-group pretest-posttest design, though we recognize that the "short" time series is also different from longer time series such as those discussed by Cook and Campbell (1979).

In a preliminary analysis of responses to the SDQ III, a 13 (self-concept factors) by 3 (times) within-subjects ANOVA was conducted for the 348 subjects who had completed responses at all three times with the commercially available MANOVA routine from the SPSS package (Hull & Nie, 1981). A 13 by-2 ANOVA was also conducted

to examine shifts in responses during the time 2 to time 3 interval. Based upon the results of these analyses, more specific comparisons were made for particular self-concept factors, as described in the Results section. A paired t-test was also used to test for the statistical significance of differences in responses to the LOC from time 2 to time 3. A more detailed examination of the program impact was made in further analyses of difference scores representing measures collected at time 2 and time 3, and incorporating the Program Director's judgments of the relevance of each of the 13 SDQ III factors to the program.

Because of the large number of statistical tests that were conducted, significance testing was done at the 0.01 level, and also at the even more conservative 0.001 level. However, partially because of the large sample size, differences significant at 0.01 were usually significant at 0.001. Since each of the self-concept scores was scaled to have a standard deviation of 10 across all responses, mean differences divided by 10 provide a measure of the change in self-concept scores in standard deviation units.

Results and Discussion

Psychometric Properties of the SDQ III and the LOC Scale.

The SDQ III is designed to measure 13 facets of self-concept, and four factor analyses were performed on responses from the instrument: separate analyses on the responses from each time and then a combined analysis across all three sets of responses. In each of the four factor analyses, the 13 SDQ III factors were clearly identified (see Table 1). Factor loadings for variables designed to measure each factor -- target loadings -- were high (median = 0.72), while nontarget loadings were small (median = 0.02). In each of the four factor analyses the smallest target loading was substantially larger than the largest nontarget loading. Correlations among the factors were modest, ranging from -.07 to 0.39 (median = 0.10). These findings demonstrate that SDQ III factors identified in other settings generalize to the responses by Outward Bound participants, and that participants' self-concepts in different areas are quite distinct.

Insert Table 1 About Here

Coefficient alpha estimates of reliability were determined for each of the SDQ III scales for times 1, 2 and 3. The coefficient alphas (Table 2) were consistently in the 0.80's and 0.90's (median alpha = 0.90 at each time) for all but the Honesty factor which also had lower reliability estimates in other SDQ III studies. Stability

coefficients from time 1 to time 2 were also in the 0.80's and 0.90's (median = .87) except for Honesty (See Table 2), and were only marginally less than the internal consistency estimates. Stability coefficients for time 1 to time 3 (median = 0.84), and for time 2 to time 3 (median = 0.83), where time 3 was after the intervention, were slightly lower; this may reflect the impact of the intervention. These findings demonstrate that each of the SDQ III scales is internally consistent, and stable over time.

Insert Table 2 About Here

The internal consistency estimates for the LOC scale are 0.71 and 0.78 for times 2 and 3 (the LOC was not administered at time 1), and were similar to results in previous research with the instrument reported by Rotter. The stability coefficient from time 2 to time 3 was 0.68 (see Table 2).

Program Impact: Change in Responses Before and After The Intervention.

Responses across all 13 SDQ III factors differed substantially with the time of testing ($F(2,696) = 136$, $p < 0.001$), but these differences due to time of testing varied significantly with the specific SDQ III factor ($F(24,8352) = 15.2$, $p < 0.001$). Mean responses to different scales were then examined separately since the time-by-scale interaction was significant. Inspection of these means (see Table 3) indicates that responses for the separate scales generally show little change or a small drop from time 1 to time 2, and then a sharp increase from time 2 to time 3. A time 1 to time 2 decline is statistically significant ($p < 0.01$) for 5 of the 13 scales, while there is a small but significant increase for one scale. The time 2 to time 3 increase is statistically significant and positive for all 13 scales, but the size of the increase varies with the scale. Inspection of the means in Table 3 also indicates that there is a significant shift in responses to the Rotter LOC scale which represents a change towards more internal responses.

Insert Table 3 About Here

The small decrease in some of the self-concept scales from time 1 to time 2 probably reflects the different settings in which the instruments were completed, and the proximity of a challenging or stressful experience. At time 1 the instrument was completed by participants in the nonthreatening environment of their own homes, about a month before the start of the course. At time 2 the instrument was completed in a new environment on the first day of the program. (In a popular magazine account one participant

described his first day: "We were all apprehensive...we knew nothing but (possibly) stretched tales about the limits to which we'd be pushed in activities we had not experienced before...we were all scared that we might be found wanting.") Consistent with this interpretation the largest time 1 to time 2 decline was for self-concept of Physical Abilities which is probably the most stressful aspect of the program. Particularly for the Physical Abilities scale, the positive impact of the program was based upon the time 2 to time 3 comparison, but there was no significant change between responses at times 1 and 3. The design of the study, the fact that time 2 and time 3 measures were administered under similar circumstances and in the same physical setting while time 1 measures were not, and the interpretation of the time 1 to time 2 shift presented above, all dictate that the time 2 to time 3 interval should be used to evaluate the program effects. Nevertheless, the shifts, particularly in the Physical Ability scale must be interpreted cautiously.

An important feature of this study is that the Outward Bound program was actually experienced by 27 reasonably distinct groups of participants which operated in different locations and at different times during the year. This multiple replication of the time-series design provides a test for idiosyncratic, time-related occurrences that are a threat to the validity of conclusions based upon a single time-series (see Cook & Campbell, 1979). In order to examine the generality of the program effects across the groups, time 2 to time 3 difference scores were determined for responses for each of the SDO III factors, and differences between the 27 groups were compared for each of the 13 scales. In only two of the 13 one-way ANOVAs did differences between the 27 groups reach statistical significance at $p < 0.05$, and none of the differences was statistically significant at $p < 0.01$. (One of the significant differences was for Relations With the Opposite Sex and this was expected since one-third of the groups were mixed-sex and the rest were single-sex -- see further discussion below). These findings indicate that effects of the Outward Bound Program were consistent across the groups and strongly support the generality of the findings.

The Lack of a No-Treatment Control Group. Changes in SDO III responses were consistent with the Outward Bound goals and points to the program's ability to enhance on self-concept. The relative lack of shift, or slight decline in self concepts over the time 1 to time

2 control interval, suggests that the time 2 to time 3 increases were not due to regression effects, spontaneous remission, and other threats to the validity of findings that are typically assessed in experimental designs with No Treatment control groups. In support of these conclusion we propose that a multiple replication time-series design is nearly as effective as an experimental design with a No Treatment control. This position is most defensible in a study such as this where shifts in the criterion measure during the pretest-posttest interval are unlikely without any intervention. When the design of the study, the selection of subjects, or the nature of the criterion make it likely that there will be a systematic shift without an intervention, this position is more dubious -- particularly when the direction of the likely shift is the same as the intended intervention effect. Even when the criterion measure is likely to change without any intervention, more sophisticated time-series designs, where there are a sufficient number of data points before the intervention to establish the trend in the shifts and/or the point at which the intervention is introduced is systematically varied, may be viable (see Cook & Campbell, 1979; Smith, et al., 1975).

Post Group Euphoria and Program Impact.

At the end of intensive group experiences, participants typically experience elation or good feelings that will be referred to here as post group euphoria (PGE). The concern here is not whether a PGE exists, or whether it is good or bad. Rather, the critical question is whether other measures are biased by a PGE so that they do not validly reflect the impact of the intervention -- particularly measures based upon subjects' self-reports. Conceptually, a PGE bias may be like a Hawthorne Effect, a placebo effect, a halo effect, an experimenter acquiescence effect in which subjects try to make the program look good, or even a cognitive dissonance effect where the time, money and effort expended by participants influences their ratings (these biases are not conceptually distinct and there may be considerable overlap among them). The possibility of a PGE bias is made more plausible by the fact that there were significant changes in all 13 SDQ III scales, even though the scales differ substantially in their relevance to the expressed goals of the program. It could be argued that the higher ratings on all the SDQ III scales are a consequence of a single global bias or halo effect produced by the PGE rather than by the intervention. Thus, A PGE bias, if it exists, represents a

source of invalidity in the self-concept responses.

There are two separate issues related to a possible PGE bias: first, whether the inferred short-term effects of the program and the self-concept responses are valid or whether they are biased by PGE; and second, whether the program effects are maintained over time, based on the assumption that a PGE is short-term. It is important to realize that the short-term gains may be valid, even if they are not maintained. For example, in a weight-loss program, if participants lose significant amounts of weight during the course of the program, the weight loss is "real" even if they eventually regain the lost weight. (If subjects' self-perceptions of their "thinness" were used to evaluate the intervention, then changes in self-reports might be due to a PGE bias). Similarly, the physical endurance and stamina of virtually every Outward Bound participant was substantially increased by the end of the program, whether or not the new level of fitness was maintained. While both issues are important, the focus of the present discussion is on the validity issue. Rephrased in terms of intervention research, the question becomes whether the Outward Bound program actually affects multiple dimensions of self-concept, or whether the apparent short-term gains are an artifact of a PGE. In terms of self-concept research, the question is one of the validity of the self-concept responses in intervention research.

A PGE bias is an important threat to the validity of any conclusions based on self-report data, and the time-series design provides little control for such an effect. Furthermore, a potential PGE bias cannot be tested with a randomly assigned No Treatment control group, and so that design may also be weak for examining intervention effects where a PGE bias is likely. Randomly assigned placebo control groups are sometimes used to explore such counter-explanations. Traditionally, a placebo is a neutral or null stimulus given to the subject as if it were the active treatment, thus allowing the researcher to separate treatment effects from psychological reactions produced by the placebo. Placebo controls are particularly effective when the actual intervention effects and subjects' psychological reactions to the intervention are conceptually distinct, as may be the case in medical research. However, placebo control groups are less viable in studies such as here where: a) the dependent variables, such as self-concept and LOC, are based upon subjects' psychological reactions; and b) it is

unlikely that a placebo could capture the psychological experience of the Outward Bound Program and still not be related to the Outward Bound goals which are tested in this study (i.e., no "Outward Bound-like" intervention is likely to be neutral with respect to its effect on self-concept and LOC). However, additional tests, based upon a more detailed analysis of the time 2 to time 3 difference scores, can be used to examine the viability of a PGE bias. In each of these analyses it is assumed that if the change scores for conceptually distinct SDQ III scales are similar in magnitude and/or highly correlated then there is support for a PGE bias, but also that if differences for the various measures differ in size and/or are relatively independent then support for a PGE bias is tenuous.

SDQ III Difference Scores. In the first set of analyses of the SDQ III difference scores, their reliability and dimensionality were examined. The logic employed here is that if a PGE bias exists, reliable changes in responses from time 2 to time 3 will reflect primarily this one factor and will not be specific to different areas of self-concept (see Marsh, Smith, Butler & Barnes, 1983 for further discussion). First, difference scores for each of the 10 (or 12) items representing the SDQ III scales were used to estimate the internal consistency of each mean scale difference. These reliability estimates, though lower than those presented in Table 2, were substantial, ranging from 0.52 to 0.82 (Md = 0.67). Second, difference scores between time 2 and time 3 were determined for each item pair as described earlier, and were factor analyzed. A total of 21 eigenvalues were greater than 1.0, and the first factor accounted for only 13% of the total variance. In a 13-factor solution, nearly all of the 13 SDQ III factors could be interpreted, though the solution was not nearly as clear as for the analyses summarized in Table 1.

Finally, differences in the SDQ III factor scores were correlated with each other and with differences in the Rotter LOC scale. The size of the correlations among the SDQ III factor differences was modest, ranging from -0.05 to 0.33 (md r = 0.10), and a majority did not reach statistical significance. Despite the small size of these correlations, the pattern of correlations was similar to the pattern of correlations among SDQ III factors at any one time. Differences in the Rotter LOC scale were nearly uncorrelated with differences in the SDQ III factors, and only its correlation with differences in the General-Self scale (-0.19) was

statistically significant.

The analyses of difference scores show that time 2 to time 3 changes in the SDQ III responses are reliable and multidimensional, and that the dimensions underlying the change scores are similar to those representing the SDQ III at times 2 and 3. The results clearly indicate that shifts in ratings between time 2 and time 3 are not unidimensional, and that shifts in self-concept are specific to particular dimensions of self-concept. However,the technical difficulties involved in the analysis and interpretation of difference scores dictate that these results must be interpreted cautiously (see Cronbach & Furby, 1970; Marsh, Smith, Barnes & Butler, 1983). Nevertheless, the findings argue against the existence of a strong unidimensional halo effect such as a PGE bias as the cause of the time 2 to time 3 shifts.

The Magnitude of Change in Different Scales Before the first data collection, the Director of the Outward Bound program was asked to indicate his perception of the relevance of each of the SDQ III scales to the goals of the program. While none was absolutely irrelevant, four -- Religion, Mathematics, Academic, and Relations With Opposite Sex -- were indicated to be clearly less relevant. In this sense, the less relevant SDQ III scales serve as a control for a PGE bias. If shifts in the less relevant goals are significant and approach the size of the more relevant goals, then there may be support for a PGE bias, though other explanations would be viable.

Time 2 to time 3 changes were statistically significant for all 13 SDQ III dimensions (Table 3), but the magnitude of the differences varied. This speculation was confirmed by the results of an additional ANOVA conducted on responses to the 13 SDQ III factors. Across all scales the shift was statistically significant ($F(1,351) = 261.6$, $p < 0.001$), but the size of the shift varied for the different scales as indicated by the time-by-scale interaction ($F(12,4212) = 12.2$, $p < 0.001$). The mean time 2 to time 3 difference score for the four "less relevant" scales was less than half that for the mean of the other nine scales (see Table 3), and the difference in changes for the two sets of scales was highly significant ($t(351) = 7.77$, $p < 0.001$). These findings demonstrate that the size of the shifts in self-concepts varied substantially for different scales, and that the differences were larger for those scales that were a priori chosen to be most relevant to the goals of the program. These findings offer further support against the existence of a PGE bias and

support for the program as having achieved it aims.

Mixed-sex and Single-sex Groups. When evaluating the relevance of the different SDQ III scales, the Outward Bound Program Director indicated that Relations With the Opposite Sex was often not relevant to the Outward Bound program because group members were typically of the same sex, but that this scale would be relevant for mixed-sex groups. As indicated earlier, 9 of the 27 groups were mixed-sex, consisting of between 31% and 50% females. Results from the single-sex and mixed-sex groups were compared on time 2 to time 3 differences for each of the 13 SDQ III scales. While the results failed to reach statistical significance for 12 of the 13 scales (all p's > .2), mixed-sex groups experienced a significantly larger improvement in Relations With the Opposite Sex scale ($t(351) = 4.5$, $p < 0.001$). These findings indicate that changes in self-concept experienced in single-sex and mixed-sex groups did not vary except for the one scale that was chosen a priori to be most relevant to this group difference -- Relations With the Opposite Sex. Again these findings argue against the effect of a PGE bias.

Summary of Discussion of A Possible Post Group Euphoria Bias. A variety of analyses were conducted to test a counter-interpretation of the time 2 to time 3 shift as a single, global halo effect such as a PGE bias. The shifts were clearly multidimensional, and shifts in different areas of self-concept were nearly independent of each other and of shifts in LOC. The size of the shifts varied significantly, and significantly larger shifts were observed in the SDQ III dimensions chosen a priori to be more relevant to the Outward Bound goals. Finally, single-sex and mixed-sex groups differed only in terms of shifts in Relations With the Opposite Sex, but not in terms of the other SDQ III factors. These findings demonstrate that the time 2 to time 3 shifts were quite specific, and consistent with a priori predictions based on program goals and group characteristics. It is still possible that such a bias had some effect in addition to the systematic effects produced by the Outward Bound program, or that specific PGE biases varied systematically with the content of each SDQ scale. Nevertheless, the systematic pattern of results argues against the existence of a strong, global PGE bias.

A PGE bias represents a serious threat to the validity of intervention studies that employ self-report criteria, and to the validity of self-concept responses in such research. A randomly

assigned No Treatment control group offers no safeguard against such an effect, and placebo control groups may be inappropriate. Thus, the further exploration of alternative tests of PGE effects is important. A particularly powerful test of a PGE bias would be the inclusion of "control" criterion measures that are likely to be impacted by PGE biases but that are unrelated to the hypothesized effect of the intervention. While such approaches are rare in experimental research and intervention studies, they are frequently used in studies designed to test the construct validity of responses to psychological measures; the logic of the approach is similar to that used in multitrait-multimethod analyses. In intervention research, threats to the interpretation of intervention effects are potential sources of invalidity in the measurement instruments, and support for the validity of the measurement instruments provides support for the interpretation of the intervention effects. Particularly in true experimental studies, this inexorable link between measurement issues and the valid interpretation of the results is given insufficient recognition.

Summary and Implications

This study supports the Outward Bound Program as an effective intervention for changing multiple dimensions of self-concept and, perhaps, locus of control. The findings, taken together with previous research, also support the validity of interpretations based upon responses to the SDQ III, and its usefulness as a criterion measure for intervention studies. Psychometric properties of the SDQ-III -- reliability, factor structure, and correlations among different dimensions -- were consistent at times 1, 2 and 3, and were consistent with earlier research based upon student responses. In addition, stability coefficients over the one-month time 1 to time 2 interval were very high. Finally, a powerful intervention specifically designed to affect self-concept was shown to significantly influence responses to the SDQ III, and the largest effects were observed with the dimensions chosen a priori to be more relevant to the goals of the program. Thus, self-concept was relatively stable, but changes did take place that were reliable, specific to particular dimensions of self-concept, and systematically related to the goals of the intervention.

Theoretical issues of importance in this study were the stability and systematic change in multiple dimensions of self-concept, the ability of effective interventions to enhance self-

concept, and the methodology of such research. Good longitudinal research in self-concept is rare (but see Bachman & O'Malley, 1977; Dusek & Flaherty, 1981; Maruyama, Rubin & Kingsbury, 1981), as are successful intervention studies that use well-designed self-concept measures (see Rosenberg, 1979; Scheirer & Kraut, 1979; Wiley, 1979), and studies that assess systematic changes in multiple dimensions of self-concept are even rarer. Hence, the findings of this study are important in that they provided information about systematic change and stability in multidimensional self-concepts, and demonstrated that self-concept can be changed through effective intervention. From this perspective a critical concern is to identify the characteristics that are important for effective intervention in self-concept research.

Three characteristics of the present study were particularly important to the demonstration that self-concept can be enhanced. First, the nature of the intervention, the 26-day residential Outward Bound program, was particularly powerful. Second, the sample size was sufficiently large so that observed changes in self-concept could be differentiated from random error. (Also, the use of subjects as their own control strengthened the power of the statistical tests). Third, the SDQ III assessed multiple dimensions of self-concept that were clearly distinguishable and some of these facets were particularly relevant to the goals of the intervention. No matter how strong the intervention, and how powerful the statistical analysis, intervention effects are likely to be negligible if they are assessed with poorly designed or irrelevant measures. Most research has relied on an ill-defined measure of overall or total self-concept, rather than specific facets of self-concept that accurately reflect the goals of the intervention. In future research we strongly recommend that these three characteristics, along with appropriate controls for PGE biases, be incorporated into the design.

REFERENCES

Bachman, J. G., & O'Malley, P. M. (1977). Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. Journal of Personality and Social Psychology, 35, 365-380.

Boersma, F.J. & Chapman, J.W. (1979). Student's Perception of Ability Scale Manual. University of Alberta. Edmonton, Canada.

Burns, R. B. (1979). The self-concept: Theory, measurement, development and behaviour. London: Longman.

Cook, T. D. & Campbell, D. T. (1979). Quasi-experimentation: Design & analysis issues for field settings. Chicago: Rand McNally.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change" -- Or should we? Psychological Bulletin, 74, 68-80.

Dusek, J. B., & Flaherty, J. F. (1981). The development of self-concept during the adolescent years. Monographs of the Society for Research in Child Development, 46(4, Serial No. 191).

Fleming, J. S. & Courtney, B. E. (1984). The dimensionality of self-esteem: II: Hierarchical facet model for revised measurement scales. Journal of Personality and Social Psychology, 46, 404-421.

Godfrey, R. (1974). A review of research and evaluation literature on Outward Bound and related educational programs. Denver, CO: Colorado Outward Bound School.

Harter, S. (1982). The Perceived Competence Scale for Children. Child Development, 53,87-97.

Hull, C. H., & Nie, N. H. (1981). SPSS Update 7-9. New York: McGraw-Hill.

Lefcourt, H. M. (1976). Locus of control. Hillsdale, NJ: Erlbaum Press.

MacDonald A. P. (1973). Internal-external locus of control. In J. Robinson & P. Shaver (Eds.), Measures of Social Psychological Attitudes (pp. 169-243). Ann Arbor, MI: Institute for Social Research.

Marsh, H. W. (1984). Verbal and math self-concepts: An internal/external frame of reference model. (In Review) Department of Education, The University of Sydney, Australia.

Marsh, H. W., Barnes, J., Cairns, L., & Tidman, M. (1984). The Self Description Questionnaire (SDQ): Age effects in the structure and level of self-concept for preadolescent children. Journal of Educational Psychology, 75, 940-956.

Marsh, H. W., Barnes, J. & Hocevar, D. (in press). Self-other

agreement on multidimensional self-concept ratings: Factor analysis & multitrait-multimethod analysis. Journal of Personality and Social Psychology.

Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III (SDQ III): The construct validity of multidimensional self-concept ratings by late-adolescents. Journal of Educational Measurement, 21, 153-174.

Marsh, H. W. & Parker, J. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? Journal of Personality and Social Psychology, 47, 213-231.

Marsh, H. W., Parker, J., & Barnes, J. (in press). Multidimensional adolescent self-concepts: Their relationship to age, sex and academic measures. American Educational Research Journal.

Marsh, H. W., Parker, J. W., & Smith, I. D. (1983). Preadolescent self-concept: Its relation to self-concept as inferred by teachers and to academic ability. British Journal of Educational Psychology, 53, 60-78.

Marsh, H. W., Relich, J. D. & Smith, I. D. (1983). Self-concept: The construct validity of interpretations based upon the SDQ. Journal of Personality and Social Psychology, 45, 173-187.

Marsh, H. W. and Shavelson, R. J. (1984). Self concept: Its multifaceted, hierarchical structure. Department of Education, University of Sydney, Australia. (in review)

Marsh, H. W., Smith, I. D. & Barnes, J. (1983). Multitrait-multimethod analyses of the Self Description Questionnaire: Student-teacher agreement on multidimensional ratings of student self-concept. American Educational Research Journal, 20, 333-357.

Marsh, H. W., Smith, I. D. & Barnes, J. (in press). Multidimensional self-concepts: Relationships with inferred self-concepts and academic achievement. **Australian Journal of Psychology** (also see ERIC ED 242 786).

Marsh, H. W., Smith, I. D., Barnes, J. & Butler, S. (1983). Self-concept: Reliability, dimensionality, validity, and the measurement of change. Journal of **Educational Psychology, 75, 772-790.**

Maruyama, G., Rubin, R. A., & Kingsbury, G. G. (1981). Self-esteem and educational achievement: Independent constructs with a common cause? Journal of Personality and Social Psychology, 40, 962-975.

Nie, N. H., Hull, C. H., **Jenkins, J. G. Steinbrenner, K. & Bent,** D. H. (1975). **Statistical Package for the Social Sciences. New**

York: McGraw-Hill.

Richards, G. E. (1977). Some educational implications & contributions of Outward Bound. Sydney: Australian Outward Bound Foundation.

Rosenberg, M. (1979). Conceiving the self. New York: Basic Books.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 80, No. 1 (Whole No. 609).

Rotter, J. B., Chance, J. E., Phares, E. J. (1972). Applications of a Social Learning Theory of Personality. New York: Holt, Rinehart & Winston.

Scheirer, M. A., & Kraut, R. E. (1979). Increasing educational achievement via self-concept change. Review of Educational Research, 49, 131-150.

Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Validation of construct interpretations. Review of Educational Research, 46, 407-441.

Shavelson, R. J., & Marsh, H. W. (In Press). On the structure of self-concept. In R. Schwarzer (Ed.), Anxiety and cognitions. N. J. Lawrence Earlbaum.

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

Smith, M. L., Gabriel, R., Schott, J., & Padia, W. L. (1975). Evaluation of the effects of Outward Bound. Boulder, CO: Bureau of Educational Field Services, University of Colorado.

Soares, L. M., & Soares, A. T. (1982). Convergence and discrimination in academic self-concepts. Paper presented at the 20th Congress of the International Association of Applied Psychology, Edinburgh, Scotland, July, 1982.

Welles, L. E., & Marwell, G. (1976). Self-esteem: Its conceptualization and measurement. Beverly Hills, Calif.: Sage Publications.

Wylie, R. C. (1974). The self-concept (Rev. ed., Vol. 1) Lincoln: University of Nebraska Press.

Wylie, R. C. (1979). The self-concept (Vol. 2) Lincoln: University of Nebraska Press.

Table 1

Summary Statistics For Target Loadings, Nontarget Loadings, and Factor Correlations in the Four Factor Analyses

		I	II	III	IV
Target Loadings (68 factor loadings)	Highest	.94	.95	.95	.95
	Lowest	.40	.44	.36	.49
	Median	.72	.74	.72	.70
	% > 0.30	100%	100%	100%	100%
Nontarget Loadings (716 factor loadings)	Highest	.21	.26	.21	.20
	Lowest	-.14	-.14	-.14	-.13
	Median	.02	.02	.02	.02
	% > 0.30	0%	0%	0%	0%
Factor Correlations (78 factor correlations)	Highest	.36	.38	.38	.39
	Lowest	-.07	-.04	-.05	-.05
	Median	.10	.11	.11	.11
	% > 0.30	4%	5%	4%	5%

I = Time 1 responses (N=357)

II = Time 2 responses (N= 358)

III = Time 3 responses (N = 355)

IV = Combined set of responses from times 1, 2 and 3 (N= 1,070)

Note: Target loadings are the factor loadings of the 68 item-pairs on the factor each is designed to measure, while all other factor loadings are Nontarget loadings. Factor Correlations are the factor pattern correlations among the 13 oblique correlations identified in each analysis. The "% > 0.30" refers to the percentage of coefficients which are greater than 0.30.

TABLE 2

Internal Consistency & Stability of SDQ III and LOC Scales

	Internal Consistency			Stability Coefficients		
SDQ Scales	Time 1	Time 2	Time 3	T1/T2	T1/T3	T2/T3
Mathematics	.94	.95	.95	.93	.92	.94
Verbal	.85	.86	.88	.90	.85	.89
Academic	.92	.93	.94	.87	.86	.89
Problem Solving	.84	.83	.86	.85	.80	.83
Physical Abilities	.92	.93	.92	.87	.86	.83
Appearance	.88	.89	.88	.84	.79	.82
Same Sex Peers	.88	.90	.88	.85	.77	.81
Opposite Sex Peers	.91	.92	.92	.88	.84	.89
Parents	.89	.89	.90	.88	.87	.89
Religion	.94	.95	.95	.94	.92	.93
Honesty	.72	.77	.77	.77	.68	.75
Emotional	.90	.90	.89	.84	.76	.81
General	.94	.94	.93	.86	.66	.73
Median coefficient	.90	.90	.90	.87	.84	.83
LOC Scale						
Total Score	---	.71	.78	---	---	.68

NOTE: Internal consistency coefficients are coefficient alpha estimates of reliability. Stability coefficients are the correlations between responses by the same individual at two points in time.

TABLE 3

Means, Standard Deviations, and Mean Comparisons

SDQ Scales	Means (& SD's) for: T1	T2	T3	t-tests of Significance for Selected Pair-wise Comparisons: T1-T2	T2-T3
Mathematics[a]	49.8 (10.3)	49.8 (10.0)	50.4 (9.7)	0.15	3.63*
Verbal	50.0 (10.0)	49.3 (9.7)	50.8 (10.3)	-3.20**	5.88**
Academic[a]	50.1 (10.0)	49.3 (9.9)	50.5 (10.1)	-3.10*	4.66**
Problem Solving	50.4 (10.2)	48.9 (9.3)	50.7 (10.3)	-5.60**	5.61**
Physical Abilities	51.3 (9.9)	48.0 (10.0)	50.7 (9.8)	-11.60**	7.95**
Appearance	49.6 (10.4)	48.3 (9.8)	52.2 (9.5)	-4.30**	12.82**
Same Sex Peers	49.5 (10.7)	49.3 (9.8)	51.3 (9.4)	-0.81	6.36**
Opposite Sex Peers[a]	49.4 (10.4)	49.6 (9.7)	51.0 (9.8)	0.62	5.36**
Parents	49.8 (10.3)	49.0 (9.7)	51.3 (9.9)	-2.47	9.01**
Religion[a]	49.8 (9.9)	49.6 (9.7)	50.6 (10.4)	-0.98	4.39**
Honesty	49.5 (9.8)	48.7 (10.2)	51.9 (9.7)	-2.17	8.39**
Emotional	48.6 (10.7)	49.6 (9.7)	51.8 (9.3)	3.05*	7.10**
General	48.7 (10.5)	49.1 (10.1)	52.2 (9.0)	1.26	8.59**
LOC Scale					
Total Score[b]	---	10.5 (3.9)	8.8 (4.3)	---	-9.43**

* p < .01, ** p < .001

a -- The SDQ III scales chosen a priori to be less relevant to Outward Bound aims by the program Director.

b -- The LOC scale is scored in the external direction so that higher scores represent a more external LOC.

NOTE: For the SDQ scales, scores represent factor scores which have been standardized to have Mean = 50 and SD = 10 across responses from time 1, time 2 and time 3. Thus, the mean difference between responses at time 2 and time 3, divided by 10, provides an estimate of changes in self-concept in standard deviation units. Scores for the Rotter LOC scale are raw responses. Responses to the 13 SDQ III scales were first analysed with 3 (time) x 13 (scale) ANOVA, and those findings indicated that the effect of time was significant but interacted significantly with the particular scale. This analysis was followed up with a series of pair-wise comparisons as presented in this table.