

Proceedings

5th Annual Research and Evaluation of Adventure Programming (REAP) 2009
Symposium: [The Evidence-Based Ingredients of Adventure Programming](#)

March 18-20, 2009
Embassy Suites Hotel
Atlanta, GA

TABLE OF CONTENTS

Speaker	Presentation Information
Dennis Embry, PhD President/CEO, PAXIS Institute	REAPing Evidence-based Kernels
Aleta Meyer, PhD Program Officer, National Institute on Drug Abuse Prevention – Research Branch	Experiential Education and Public Health: The Potential to MOVE America Toward Health
Judy Kruger, PhD Epidemiologist, Center for Disease Control and Prevention, Physical Activity and Health Branch	Integrating the 2008 Physical Activity Guidelines for Americans and the Guide to Community Preventive Services into Adventure Programs
Scott Hartl Director of Strategic Programs and Partnerships, Expeditionary Learning Schools Outward Bound	Expanding What Counts as Success in Expeditionary Learning Schools: Measuring Character and Real-World Project Work along with Academic Achievement
Aleta Meyer, PhD Program Officer, National Institute on Drug Abuse Prevention – Research Branch	Seeking Funding Through a Small Business Approach
Lee Gillis, PhD Professor of Psychology, Georgia College & State University	Educational and Therapeutic Meta-Analysis Benchmarks
Elizabeth Speelman Assistant Director of Program Development, Youth Learning Institute, Clemson University	
Maurie Lung, LMHC/LMFT Director of Performance Improvement, Eckerd Youth Alternatives, Inc. Greg Holeman Chief Quality Officer, Eckerd Youth Alternatives, Inc.	Balancing Evaluation and Program Needs
Jon Frankel Research and Development Manager, Outward Bound USA Alan Ewert, PhD Professor, Endowed Chair in Outdoor Leadership, Indiana University	Evaluation and Collaboration: Examining Outward Bound's Research Initiatives
Alan Ewert, PhD Professor, Endowed Chair in Outdoor Leadership, Indiana University	Evidence-Based Research: Practice and Learnings from Other Fields
Lee Gillis, PhD Professor of Psychology, Georgia College & State University Aaron Nicholson Executive Director, Project Adventure South East Kim Boykin Clinical Director	The Effectiveness of Project Adventure's Behavior Management Programs for Male Offenders in Residential Treatment
Matt Claybaugh, PhD President/CEO, Marimed Foundation With Scott Allred, Gane Danes, and Jamie Burke	Phase II: After the Evidence, Rolling up the Sleeves

REAPing Evidence-based Kernels to Better the World

Presented at the Research and Evaluation of Adventure Programming 2009 Symposium: Atlanta, GA

Dennis Embry

Article by Dee Wilson, Lecturer
University of Washington School of Social Work
(206) 221-4278
wilsod@u.washington.edu
Used by permission

"Evidenced Based Kernels: Fundamental Units of Behavioral Influence," by Dennis D. Embry and Anthony Biglan, Clinical Child and Family Psychology Review, 11, 75-113, 2008.

This is one of the most important articles in prevention research published in recent years. Embry and Biglan provide "an analysis of fundamental units of behavioral influence that underlie effective prevention and treatment." These authors call these units "kernels" which they define as "fundamental units of behavior influence in the sense that deleting any component ... would render it inert." According to Embry and Biglan, kernels are the common ingredients which make evidence based programs work. They write that "a kernel is like a seed that contains central information for growth or change." Embry and Biglan identify 52 kernels "shown empirically to affect a behavior;" they write that "our criteria for empirical support include RCTs (randomized trials) and interrupted time – series designs in which a procedure's impact is evaluated on a repeated measure of target behavior." Their view is that time series designs embody an inductive, incremental, bottom – up strategy "to build effective behavior – influence practices." They assert that "Kernels are of necessity simple steps targeting a behavior one can easily measure repeatedly."

Embry and Biglan offer a taxonomy of kernels: (a) many kernels involve praise and/ or rewards or (less often) negative consequences ; social approval immediately applied in response to desired behavior is powerful with all age groups (b) some kernels involve the use of signals to elicit behavior which then may bring rewards, for example the use of stop lights in school settings or traffic settings (c) kernels may influence behavior through relational framing in which a desired behavior is paired with status, belonging, protection, safety, pleasure or identity, as in advertising (d) some kernels alter behavior through physiological reactions, for example, the ingestion of omega – 3 fatty acids, massage or deep breathing.

Consider the following list of kernels from Embry and Biglan's article: verbal praise, peer to peer written praise, use of tokens or reward tickets, team competition, time out, taxes on consumptive behaviors, non - verbal transition cues, cooperative play, errorless discrimination training, paragraph shrinking, public commitment, "Us" vs. "Them" role framing, motivational interviewing (MI), massage, use of omega 3 fatty acids, rough and tumble play, aerobic play, nasal breathing. Some of these practices e.g., praise, can be utilized in most social interactions; others are mainly used in educational settings while motivational interviewing involves a set of skills which approximate a program. Embry and Biglan's list of kernels seems ad hoc and a bit arbitrary despite their useful taxonomy. Other kernels which could arguably be added to the list without a lengthy review of evidence based programs include anti - depressants, positive self talk, conferring of status, sacred symbols, empathetic listening (the kernel in MI?), pro- social roles, repetition in training programs or in rhetoric, intermittent reinforcement, mirroring of social behavior and obstacles to romance. Obstacles to romance as a means of intensifying emotional reactions may lack RCTs but who can argue with hundreds of years of western literature?

Embry and Biglan comment that current thinking about prevention and treatment "assumes that we will identify empirically supported programs and, to a lesser extent, policies, and will disseminate them widely and effectively." These authors discuss several limitations to this strategy: (1) "First, it is difficult to implement a program's efficacy widely with fidelity or effectiveness," they assert. It is difficult to assure that programs are implemented with fidelity, and even when this obstacle is overcome, "the obtained effectiveness results do not replicate the efficacy trials," they write. Embry and Biglan believe that "sole reliance on program dissemination to affect population outcomes will have a limited impact, even with restrictive policies." (2) In addition, many problems and behaviors "that affect well being do not require lengthy or complex interventions involving consultations, workshops, training or support," in their view. They offer the example of teachers' use of response cards to increase academic achievement; this is a kernel whose dissemination does not require a manual, expensive and lengthy training program or consultation from program developers. (3) Embry and Biglan argue

that “program dissemination is unlikely to affect practices or problems that fall outside the scope of a program.” Evidence based kernels can help parents and teachers with a wide range of child behaviors for which programs do not exist or are not needed, in the view of these authors. (4) Evidence based programs (EBPs) are costly to implement. Publicly funded organizations often have difficulty implementing one or two EBPs with a high degree of fidelity, much less a whole range of programs. (5) Embry and Biglan assert that “existing programs have limited effectiveness, modest effect sizes, scalability concerns, weak generalization, difficulty with maintenance and even iatrogenic effects.” They add that “this is not to diminish the enormous progress of prevention science in the past 30 years. However, inspection of recent meta-analyses of interventions ... finds plenty of room for improving the effectiveness of our programs.” (6) One reason EBPs are difficult to disseminate is because they “do not easily meet the diffusion criteria” identified in studies of innovative practices, for example, the influence of individuals described in studies as “early adopters”. The dissemination of EBPs usually requires organizations to pay the cost of implementation. “If a school or agency lags, thousands of individuals or families in geographic areas cannot avail themselves of strategies that might prevent school failure, substance abuse, mental illness, delinquency, or other ills,” Embry and Biglan state.

These authors are not suggesting that evidence based practice can proceed without program development, testing and implementation; rather their view is that much can be done without implementing programs. They also provide a child welfare example of how kernels can be packaged to build promising programs and community interventions. For example, a community could give contacts to organizations to develop a prize bowl by recruiting substance abusing adults in jails, emergency rooms, shelters, courts and churches, giving out rewards for sobriety and recovery and monitoring results across settings. Substance abusing individuals in the program can be given Omega – 3 supplements to reduce mood disorders and aggressive behavior. Infants born to addicted mothers would be placed in kangaroo care to reduce developmental problems. Toddlers and preschoolers would receive errorless compliance training by the birth parent(s), foster parents or pre- school teachers. “Exposed preschoolers and elementary children under court petition or in special ed receive self modeling videos or digitally created storybooks for social skills and behavior at home, foster care or care settings.” Local governments and schools would promote community wide praise notes from adults to reinforce developing social competence and pro- social behavior in school age children with maltreatment histories. School age children under court structure would receive ‘red flag’ training to reduce explosive anger and aggression; and juvenile justice, emergency room and school personnel would conduct motivational interviews for youth engaged in problem behaviors. In this example, Embry and Biglan suggest a packaging of kernels and EBPs, all of which have strong empirical support, to organize a community wide intervention for methamphetamine addicted parents and their children from infancy through adolescence.

Embry and Biglan advocate for “a database repository of kernels, analogous to the human genome project ...” Embry and Biglan’s list has numerous kernels tested in classrooms and in parenting programs. The list could be easily expanded by knowledgeable therapists and scholars familiar with the EBP literature and with child abuse and neglect prevention programs.

Character, Motivation, and Engagement in Expeditionary Learning Schools: A Review of the Relevant Literature and Available Measurement Instruments

*A Summary Report for Expeditionary Learning Schools Staff and Friends**

Presented at the Research and Evaluation of Adventure Programming 2009 Symposium: Atlanta, GA

By Chris S. Hulleman, Vanderbilt University; Scott Hartl, Expeditionary Learning Schools; and Keith D. Ciani, University of Missouri-Columbia*

* Note: abbreviated version of original paper; for a full text version, please contact Scott Hartl at shartl@elschools.org.

*Special thanks go to Libby Woodfin, Mindy Spencer, Ron Berger, Tom VanWinkle, and Mark Conrad for their help with this report and to Odyssey School for the cover artwork (on the original).

INTRODUCTION

Background

Intelligence plus character, that is the true aim of education. (Martin Luther King Jr.)

Effective Expeditionary Learning schools offer compelling testimony that school success includes, but cannot be limited to, strong achievement on high stakes tests. The best Expeditionary Learning schools post strong, sometimes spectacular, results on traditional measures of student achievement such as state tests and college placement rates, but a visit will make it quickly evident that these measures alone are not adequate to document the growth that students are demonstrating in these schools.

This literature review is aimed at better understanding student growth in important outcome areas that remain largely uncounted in the current prevalent definition of school success and student achievement. Traits such as initiative and self-direction, work ethic and time management, positive attitude, effective collaboration, and the ability to create high quality work products and performances are not measured or reported by state tests, but they are essential indicators of real-world achievement and readiness for college and the demands of the 21st Century work force.

To capture the multiple aspects of student achievement that we see in our best schools, Expeditionary Learning Schools has articulated a vision of school success that is broader than test scores alone. We describe school success in three groups of synergistic indicators that collectively provide a rich and real-world vision of student achievement and readiness for college and 21st Century work. Our measures of student achievement and school success include:

Traditional achievement and engagement indicators including, criterion and norm referenced test scores, attendance, graduation rates, and success in post-secondary education.

Character, motivation, and engagement indicators that include students' habits of performance and relational character; motivation to learn as evidenced by high expectations, ability to self-manage, and connectedness to school; and student engagement with learning.

Creating and presenting high quality products and performances that demonstrate the types of competencies called for by the 21st Century Skills Movement, the TIMMS reports, and other well established standards of readiness for real-world challenges.

This abbreviated report summarizes the attributes and availability of reliable measures in the second set of these indicators, *Habits of Work and Character, Motivation to Learn, and Engagement with Learning*.

Overview

Supported by fifteen years of anecdotal evidence and observation, it is our belief that students in effective Expeditionary Learning schools demonstrate unusually strong growth in their motivation to learn, engagement in school, and habits of work and character. We have not however, developed precise language to describe these gains, nor have we developed measurement tools to test our beliefs and assumptions. This literature review aims to address these important gaps.

Theory of Change

Character, Motivation, and Engagement in Expeditionary Learning schools

As presented in Figure 1, the ELS theory of change proposes that the curriculum, pedagogical approach, and culture and climate present in effective Expeditionary Learning (EL) schools support positive growth in students' habits of work and character and their motivation to learn, and further, that growth in these areas supports a heightened level of student engagement in their own learning. Student engagement in learning is the crucial variable that then leads to student success outcomes, such as performance on standardized tests, and the development of 21st century work readiness skills and personal interest in specific academic topics. It is our belief that habits of work and character, motivation to learn, and student engagement in learning are predictive of increases in student success outcomes, as well as being valuable outcomes in themselves. We also believe that these relationships are reciprocal; that is, not only do effective EL schools increase motivation and engagement in the learning process and facilitate the development of character, but motivation and character support more effective EL schools.

Working from the top of our model, we name three attributes that are consistently present in all effective Expeditionary Learning schools: 1) *Motivating and Demanding Academic Projects and Purpose*; created by well-planned Learning Expeditions, projects and authentic audiences for student work; 2) *Active and Engaging Teaching and Learning Strategies*; well planned lessons that focus on the purposeful engagement of students with meaningful content and learning activities; and 3) *Strong and Positive Whole-School and Classroom Culture*; that focuses on building a safe, respectful, and learning-focused community.

These qualities in a school support growth in students' *Habits of Work and Character* and in their *Motivation to Learn*. We define habits of work and character in this model through two groups of attributes: *Performance Character*, which we define as the traits that enable students to perform to their potential; and *Relational Character*, which we define as the traits that enable students to relate well with others and be good community members.

Effective Expeditionary Learning schools support growth in performance and relational character by defining what these habits of character mean and look like in ways that students can understand and buy into.

We have used the terms *Performance* and *Relational Character*, which have their roots in the character education literature, to describe the habits of work and character we want to study in students for two reasons. First, these terms powerfully describe what we see in students in effective Expeditionary Learning schools as they grow in their ability to perform to their potential more consistently and they become positive participants in the school community. Second, these terms are associated with a well-established literature and an emerging group of measurement scales and instruments.

We define a student's motivation to learn through three groups of constructs as differentiated in the literature: *High Expectations* (the extent to which students believe that they can learn and attain important outcomes), *Self-Management* (the extent to which students can regulate their motivation and engagement in learning), and *Connectedness/Attachment to School* (the extent to which students feel connected to school). Each of these motivation constructs is highly descriptive of what we believe is happening regarding student motivation to learn in effective Expeditionary Learning schools. Furthermore, we believe that students' motivation to learn is positively impacted by the curriculum, pedagogy, and culture of effective Expeditionary Learning schools.

Our model separates student *Engagement with Learning* from character and motivation constructs. We make this differentiation because we believe that *Engagement with Learning* plays a particular role in connecting increases in students' *Habits of Work and Character* and *Motivation to Learn* to our indicators of *Student Success*. We believe that

growth in students' habits of work and character and motivation to learn are important outcomes in themselves and as such are worthy of effective measurement.

We define Student Success Outcomes broadly in this model. We include increased student achievement as measured by student performance on state and national tests, school grades and attendance, college acceptance and completion rates, and the quality of students' work products and performances. We also include 21st Century Work readiness as described by the Partnership for the 21st Century Skills. Although perhaps not traditionally considered an important educational outcome, *Enduring Interest in Academic Topics* is included as it is a driving force behind the decisions students make in terms of their future educational and career choices.

Literature Review

Motivation: Conceptual Review

Based on the theory of change outlined above and presented in Figure 1, our literature review focused on three general areas of student motivation relevant to the ELS approach: high expectations, self-management, and connections with school. Within each of these areas, we further differentiate and review the constructs (i.e., concepts) that are most relevant to the ELS approach. The individual motivation constructs are summarized in Table 1 and expanded upon below.

High Expectations

A Preference for Challenging Activities: A preference for challenging activities can be defined as the extent to which students prefer to work on tasks that are equal to or just beyond their abilities, as opposed to easier tasks where success is more likely. More learning and growth is expected to occur when engaged with challenging tasks as temporary failure can encourage new learning and strategy use.

Academic Self-Efficacy: Academic self-efficacy is the extent to which students believe that they can successfully perform specific academic tasks, such as learning the material and doing well on exams, and is positively associated with learning, academic achievement, and future course enrollment.

Importance of Achievement: The importance of achievement is defined as the extent to which students want to do well in school, experience pride in success, and are personally invested in the pursuit of competence. In other words, they place a high value on being competent in school activities, and their identity and sense of self is connected to their achievement (i.e., attainment value). Research has demonstrated that valuing competence and placing personal importance in doing well on an activity is one of the key ingredients in the development of intrinsic motivation.

Hope: Hope is the belief in a positive future, and has been researched as two different constructs within the motivation literature. Dispositional hope is defined as the extent to which an individual believes that they can generate successful plans to reach their goals (pathways), and sustain the level of commitment and determination necessary to achieve their goals (agency). Hope scores correlate positively and significantly with academic performance for children of all ages, controlling statistically for intelligence, self-esteem, and prior grades and standardized test scores. A similar construct to dispositional hope is situational optimism (i.e., the expectations of positive outcomes in a specific context, such as school, sports, or work).

Self-Management

Intrinsic and Extrinsic Motivation: Rather than simply focusing on *how much* motivation a student has, researchers have also focused on the *type* of motivation students have for particular activities. Two types of motivation that concern the reasons why students engage in academic activities are *intrinsic motivation* and *extrinsic motivation*. A focus on intrinsic, as opposed to extrinsic, reasons for engaging in learning activities is associated with increased effort, persistence, achievement, and higher quality learning.

Views of Intelligence and Learning Goal Orientations: Carol Dweck's (1998) work on students' *views of intelligence* has proven to be an important construct in the field of education. When students believe that their mind is a muscle, and

that their skills and abilities can be improved through effort, they are more likely to put forth effort into learning activities, persist when faced with challenge or difficulties, and perform better. Students with entity views are also more likely to adopt a *learning goal orientation*. Learning goals serve as mental frameworks that orient students to their learning activities and are associated with persistence on difficult tasks, adaptive help-seeking, cooperative learning, deeper and higher quality learning, increased self-efficacy, achievement, and interest in academic topics.

Connections and Attachment to School

Relevance to Life: Students can find meaning and value in their schoolwork for a variety of reasons. Discovering that school is relevant to their lives has been termed by various researchers as *relevance to life*, *utility value*, and *perceived instrumentality*. When students discover meaning and relevance of their coursework to their lives, they become more interested in their coursework, put forth increased effort, and perform better.

School and Classroom Belonging: School and classroom belonging is the degree to which students feel like they belong as a part of their classroom and school communities. Since learning is a social activity and occurs in a social context, it follows that students' perceptions of how they relate to one another would be associated with their goals for learning. This sense of community for students may encompass students' feelings of connection and friendship to others in the class as well as their sense of how well they "fit in" socially to the general school environment. Striving towards challenging learning objectives with members of the group enhances belongingness and connection, particularly when active learning pedagogies are used, and leads to increased academic achievement and perceptions of self-efficacy.

Interest in Specific Academic Topics: Contemporary theorists have divided interest into two components: individual interest and situational interest. Enduring (individual) interest is more trait-like and persists over time. In contrast, situational interest is more momentary and context-specific. Over time, situational interest can be developed into more enduring individual interest that can direct students academic and career choices beyond high school.

Character: Conceptual Review

A major source of literature regarding character comes from the character education movement, funded in large part by the federal government, as a part of creating comprehensive and effective models of school reform. This movement has been spearheaded more by practitioners than researchers, and has focused more on changing school culture to positively impact student outcomes and less on defining and differentiating character constructs. The focus of these projects tends to be on measuring traditional student outcomes – such as learning, achievement, and behavior – and less on how those outcomes were attained (i.e., changes in character). The major challenge of this character review was to typify the character constructs relevant to the ELS approach, despite the lack of focus on how character is defined, and which components of character are the most important in determining educational outcomes.

Components of Character: Our synthesis of the character literature as it pertains to the ELS approach precipitated three separate, but related, aspects of character: performance character, relational character, and moral character. This framework expands on Lickona and Davidson's performance and moral character dichotomy by differentiating relational character as a unique component. Performance character refers to those personal traits and skills that are needed for doing excellent work and performing to one's potential, such as effort, initiative, diligence, self-discipline, and perseverance.

Relational character refers to those personal traits and skills that are needed for interacting effectively with others in groups, teams, or communities, such as justice and fairness, honesty, social responsibility, helpfulness, productive and appropriate collaboration skills, and social competence. Moral character refers to those personal traits and skills that are needed for ethical behavior, such as caring, honesty, fairness, responsibility, and respect for self and others. Rather than being completely independent, these aspects of character are mutually supportive.

Research evidence: The research evidence linking character constructs to student outcomes comes from a diffuse array of sources. First, reviews of character education programs tend to find small, but statistically significant effects of character education programs on student achievement, and aspects of character such as socio-moral cognition, relationships with students and teachers, and pro-social behaviors and attitudes. Second, research in psychology provides some evidence that some components of character, such as self-discipline (i.e., self-control), self-concept and self-esteem, work ethic (i.e., need for achievement), effort, and curiosity and creativity are associated with positive outcomes

such as academic performance, standardized tests, work productivity, life satisfaction, happiness, and well-being. Third, recent conceptualizations of character and development of their instruments has demonstrated positive relationships between character and a limited range of outcomes, such as life satisfaction and GPA.

Fourth, research within the positive youth development literature reveals small to moderate positive correlations between psychosocial maturity and academic achievement; and between developmental assets and academic achievement, increased thriving behaviors (e.g., succeeds in school, helps others, exhibits leadership, etc.), and reduced high-risk behaviors (e.g., alcohol and drug use, sexual activity, violence, etc). A fifth source of evidence for the role of character in promoting student success outcomes comes from research on motivation and engagement. Many of the constructs we reviewed in the motivation and engagement section have considerable overlap with our conceptualization of performance character in particular.

Engagement with Learning: Conceptual Review

Learning Involvement: Ideally, educational activities will completely engage students such that they become caught up and lost in their work; in other words, they become cognitively and behaviorally involved in their learning (i.e., *learning involvement*). Learning involvement has been associated with increased intrinsic motivation, achievement, topic interest, effective teaching and learning, creativity, and peak sports performance.

Classroom Engagement: In addition to being caught up in the learning process, students can also become engaged in learning through their enjoyment of the course (*emotional engagement*), or through effort, participation, persistence, and attention (*behavioral engagement*). Active learning activities that allow students to enjoy and explore the material are likely to elicit both emotional and behavioral engagement. Both behavioral and emotional engagement have been positively correlated with intrinsic motivation, negatively correlated with extrinsic motivation, and their aggregate measure has been correlated with academic achievement, hope for the future, and perceived teacher support for learning.

Student Success Outcomes: A Brief Review

In the ELS model, student success is only partially comprised of achievement as measured by standardized tests or grade point average (GPA). Student success outcomes of the ELS approach also include work readiness skills demanded by the 21st century and the development of enduring (individual) interest in specific academic topics.

21st Century Work Readiness: There are two main frameworks of work readiness skills that we drew upon in our literature review: the Partnership for 21st Century Skills and the Secretary's Commission on Achieving Necessary Skills (SCANS). The Partnership for 21st Century Skills has outlined four pillars of 21st century skills: learning and innovation skills, life and career skills, information media and technology skills, and core subjects and 21st century themes. Learning and innovation skills, along with life and career skills, overlap most strongly with the goals of the ELS. These two pillars of the Partnership's 21st Century Skills framework overlap with two aspects of the Secretary's Commission on Achieving Necessary Skills (1991) framework: the five competences (resources, interpersonal, information, systems, and technology), and personal skills and qualities (basic skills, thinking skills, and personal qualities).

There has been remarkably little progress on developing measures of any of the four pillars. For example, the Partnership has an online resource, called Route 21 (<http://www.21stcenturyskills.org/route21/index.php>), a "one-stop shop for 21st century skills-related information, resources and community tools." As of January 16, 2009, there were 570 tools listed within the Route 21 website. Of these tools, 423 were categorized as tapping learning and innovation skills, and 407 were categorized as tapping life and career skills. However, only two of these tools were categorized as instruments that measure either learning and innovation or life and career skills. The Cisco Networking Academy has developed an instrument, the Student Engagement and Best Practices Survey, which assesses career self-efficacy, work responsibility, collaboration and teamwork, life-long learning, problem-solving confidence, motivation, academic self esteem, learning effort, active learning, and level of challenge. The Organization for Economic Co-operation and Development has developed an instrument, the PISA Self-regulated Learning Assessment, that measures cognitive and metacognitive learning strategies, motivation and interest, self-concept, and preference for learning situations.

CONCLUSION

Next Steps

In sum, there is a trade-off between precision and efficiency that needs to be considered by ELS evaluators when developing an instrument to tap students' *Habits of Work and Character*, *Motivation to Learn*, *Engagement with Learning*, and *21st Century Skills*. The aim of this report was to provide a foundation from which the importance of precision and efficiency can be adequately considered. Despite the lack of well-developed scales to measure character and its subcomponents, evaluators may be able to capitalize on the overlap between the definitions of character considered in our review and the rich body of motivation literature. Many of the motivation constructs relevant to the ELS approach (intrinsic motivation, interest in academic topics) also overlap with performance and relational character (positive attitude, love for learning, curiosity). Instead of completely 'reinventing the wheel', well-developed motivation instruments can be utilized as proxies for the related components of character. The conceptualization and measurement of 21st century work skills relevant to EL schools is even less developed than character, and thus we suggest that motivation and character instruments could be used as starting points for the development of work readiness instruments. Our 'Top 10 List' of instruments and constructs presented in Table 1 provides a solid foundation for the instrument development phase of this project. We recommend that ELS evaluators start with more measurement in a pilot testing phase, and subsequently making reductions in measurement as the most essential constructs are identified empirically. This initial research will help refine the theory of change and determine which measures are indeed most affected by the ELS approach and most predictive of student success outcomes.

Table 1. Top 10 Instruments to Measure Students' Character, Motivation, and Engagement in Expeditionary Learning Schools

Instruments	Constructs Measured	Source/Authors
1. The Hope Study	Learning Goal Orientation, Belongingness to School, Academic Engagement, Dispositional Hope	Van Ryzin et al. (2008)
2. Collective Responsibility for Excellence and Ethics	Performance and Moral Character	Khmelkov & Davidson (2008)
3. Perceptions of Utility Value and Meaningfulness	Relevance to Life	Hulleman et al. (2008); Mitchell (1993)
4. Learning Involvement Scales	Involvement	Hulleman (2008); Mitchell (1993)
5. Situational and Personal Interest Scales	Situational and Individual Interest	Harackiewicz et al. (2000); Mitchell (1993)
6. Values in Action Inventory – Youth	Performance, Relational, and Moral Character	Park & Peterson (2006)
7. Child Development Project Student Questionnaire	Relational Character	Developmental Studies Center, (2005a, 2005b)
8. Motivated Strategies for Learning Questionnaire	Self-efficacy	Pintrich et al. (1991, 1993)
9. Academic Self-Regulation Questionnaire	Intrinsic vs. Extrinsic Motivation	Ryan & Connell (1989)
10. Profiles of Student Life: Attitudes and Behaviors	Performance, Relational, and Moral Character	Benson et al. (1998) Leffert et al. (1998)

Figure 1

CHARACTER, MOTIVATION AND ENGAGEMENT IN EXPEDITIONARY LEARNING SCHOOLS AND THEIR RELATIONSHIP TO STUDENT ACHIEVEMENT AND 21ST CENTURY WORK READINESS

ATTRIBUTES OF AN EFFECTIVE EXPEDITIONARY LEARNING SCHOOL

- **Motivating and Demanding Academic Projects and Purpose**
- **Active and Engaging Teaching and Learning strategies**
- **Strong and Positive Whole-School & Classroom Culture**

HABITS OF WORK AND CHARACTER

Performance Character

Traits that enable students to perform to their potential:

Effort Initiative Perseverance Work Ethic Positive Attitude Organization Time Management

Relational Character

Traits that enable students to be good people and community members:

Justice and Fairness Respect for Others Social Responsibility Environmental Responsibility Conflict Resolution Helpfulness Communications Competence Civic Knowledge and Engagement Productive and Appropriate Collaboration

MOTIVATION TO LEARN

High Expectations

Belief in the ability to attain important outcomes:

Preference for Challenge Academic Self-Efficacy Competence Valuation Educational Aspirations and Expectations

Self-Management

Regulation of motivation and engagement in learning:

Intrinsic vs. Extrinsic Motivation View of Intelligence Fixed vs. Changeable Orientation to Mastery Desire to Develop and Improve

Connectedness and Attachment to School

Feeling connected to the people and work of school:

Classroom Community Relevance and Connection of School to Students' Lives Interest in Specific Academic Topics

ENGAGEMENT IN LEARNING

STUDENT SUCCESS OUTCOMES

Academic Achievement

Increased State and National Test Scores	Enduring Interest in Academic Topics
Increased School Grades & Attendance	Quality Student Work Products and Performances
Increased college Acceptance & Completion Rates	

21st Century Work Readiness

<u>Life and Career Skills</u>	<u>Learning and Innovation Skills</u>	<u>Information, Media, and Technology Skills</u>	<u>Core Subjects and 21st Century Themes</u>
Leadership & Responsibility Initiative & Self-Direction Social & Cross Cultural Skills Productivity & Accountability Flexibility & Adaptability	Creativity & Innovation Skills Critical Thinking & Problem Solving Skills Communication and Collaboration Skills	Information and Media Literacy	Global Awareness Financial Literacy Civic Literacy Health Literacy Environmental Literacy

Therapeutic and School-Based Meta-Analysis Benchmarks

Presented at the Research and Evaluation of Adventure Programming 2009 Symposium: Atlanta, GA

By Lee Gillis, PhD, Professor of Psychology, Georgia College & State University, lee.gillis@gcsu.edu; and Elizabeth Speelman, Assistant Director of Program Development, Youth Learning Institute, Clemson University, lizspeelman@gmail.com.

Introduction

Google searches for "challenge course" or "ropes course" reveal 405,000 and 616,000 entries respectively (retrieved November, 28, 2008). One common question associated with such use is the outcome or impact of participation in such experiences. For example, when compared to a control group, what efficacy exists for challenge course programs conducted with therapeutic intent? Or how would a researcher or practitioner determine if such programs provided a recognized "evidence base" for the field of adventure therapy? Are the predominantly school-based participants used in challenge course research truly representative of all groups? Are there more effective outcomes for either therapeutic or school-based groups?

Challenge course programming

Rohnke, Rogers, Wall, and Tait (2007) cited 1941 as the origin of challenge ropes courses (hereafter referred to as challenge courses) when obstacle courses were first used in military training. Challenge courses were introduced in the United States at the Colorado Outward Bound School in the early 1960s. Challenge courses are currently used in recreational, educational, developmental, and therapeutic settings as well as in camps, hospitals, and corporate training centers. The focus of this study is to establish initial effect size benchmarks for the use of challenge course activities with therapeutic outcomes where adventure therapy may be taking place and for school-based groups utilizing cooperative group games and challenge course activities.

Method

Studies for the original meta-analysis were selected based on the following criteria:

1. The study focused solely on challenge course activity outcomes.
2. The study was conducted between the years 1986 and 2006.
3. The study had to include control groups. Control groups received no challenge course programming, were on a waiting list for programming at a future time, or received alternative programming.
4. Outcome measures had to be recorded with sufficient data to compute effect sizes.
5. The studies had to be accessible to the authors.

Eligible studies

Of the 137 potential studies, a total of 44 studies matched the original criteria for inclusion as listed above. All of the studies were conducted in the United States. Activities were examined according to the intended purpose of the program that was being researched: recreational, educational, developmental, or therapeutic (Priest, Gass, & Gillis, 2000/2003). Fifteen studies were identified or with the intent to change patterns of behavior of the group participants (therapeutic focus). Studies were also coded according to the setting in which the research was being conducted. Twenty-five studies were identified as taking place in a school setting.

Study descriptors and outcome measures. Information was collected regarding the date of the research, population studied, activities conducted, purpose of the activities, and whether the research was published for each study. Populations were coded according to age, race, gender, and type of group (e.g. school, therapeutic). Challenge course activities were coded as only low activities or a combination of low and high, and by duration (hours). Each study was coded by the specific outcome measures being researched.

Effect Size

Calculating. Effect sizes were calculated for each quantitative measure for each study using an effect size computation program (Wilson, 2001). The resulting effect size was a mean difference (Cohen's d). For the purpose of this study, reported effect sizes represent the difference, positive or negative, between results on outcome measures for participants in the challenge course group and the control group.

Results

Standardized mean difference effect size and confidence intervals are presented in Table 1 for each of the studies. An overall effect size of 0.52 (95% CI 0.35 - 0.69) was calculated from the 15 therapeutic studies and an overall effect size of 0.25 (95% CI 0.15 - 0.36) was calculated from the 25 school-based studies. Additional information for each study includes author, year of the study, mean age, number of participants in both treatment and control groups, duration (in hours), and number of effect sizes.

In Table 2, the highest effect sizes for therapeutic focused studies were found for the following three outcomes: academic measures, self-efficacy, and family measures. Group dynamics (e.g., cohesion, effectiveness) accounted for one of the smallest effect sizes.

Table 3 presents the number of studies and effect size statistics for school-based participants by outcome constructs. Studies measuring self-esteem and self-concept account for the greatest number of studies by construct (48%) although they have one of the lowest mean effect sizes ($d = 0.17$). The highest mean effect sizes are for behavioral observations ($d = 0.41$) and group dynamics ($d = 0.66$).

Discussion

The goal of this paper is to establish benchmarks of effect sizes with confidence intervals for various aspects of challenge course studies identified as having therapeutic intent or having school-based participants. To that end it is successful. Overall an initial "line in the metaphorical sand" of outcomes has been drawn.

There are numerous limitations with this study that should be addressed. First and foremost is the small number of eligible studies in the available literature that meet the criteria for the meta-analysis. Second, the amount of missing data resulting from authors not listing basic information about their programs, while not new to this field (c.f. Gillis & Thomsen, 1996; Authors, 2008), significantly limits the generalization of findings to core aspects of challenge course programming: duration of program, age of participants, and which activities and elements were employed in the design of the experience. Practically, there exists a significant gap of knowledge in this study with no information about the amount and type of facilitation or the level of expertise of the facilitators.

Challenge course programming in therapeutic programs and with school-based participant can provide a positive change for the individual needs of each participant through the group process. Researchers should focus on improved research techniques applied to challenge course programming in group based therapy to provide greater treatment fidelity and an increased understanding of the potential of this form of experiential, activity-base practice.

References

- Authors (2008). Are challenge (ropes) courses an effective tool? A meta-analysis. *Journal of Experiential Education*, 31(2), 111-135.
- Gillis, H. L. & Thomsen, D. (1996). A Research Update (1992 - 1995) of Adventure Therapy: Challenge Activities and Ropes Courses, Wilderness Expeditions, & Residential Camping Programs. Martinsville, IN: Bradford Woods, Indiana University: Coalition for Education in the Outdoors Symposium Proceedings.
- Priest, S., Gass, M., & Gillis, L. (2000/2003). *The essential elements of facilitation*. Seattle, WA: Tarrak Publications.
- Rohnke, K., Rogers, D., Tait, C. M., & Wall, J. B. (2007). *The complete ropes course manual*. Dubuque, IA: Kendall/Hunt Publishing Company.
- Wilson, D. P. (2001). Effect size determination program. College Park: University of Maryland

Tables

Table 1.

*Study author, date, age, number of subjects, duration, setting, number of effect sizes, mean effect size, and confidence intervals for each study (**therapeutic studies/**school-based studies)*

Author	Year	Age	n (Tx)	n (Ct)	Hours	n (d)	M (d)	-95% CI	+95% CI
Aghazarian**	1997	16.5	46	49	8.0	18	0.02	-0.61	0.65
Bisson**	1998	11.0	54	56		8	2.83	2.14	3.52
Blanchard*	1994	14.8	20	20	27.0	5	0.53	-0.10	1.16
Boyle**	1986	24.7	30	30	7.0	10	0.02	-0.60	0.64
Brehey**	2000	19.2	21	18	4.0	24	0.11	-0.66	0.89
Daheim**	1999	12.7	110	86	8.0	6	0.10	-0.31	0.50
Eagle**	2000	13.0	74	64	37.0	8	0.13	-0.21	0.46
Faubel*/**	1998	13.0	41	27		6	0.74	0.24	1.24
Faulkner*	2002		54	66	3.0	6	0.89	0.36	1.42
Finkenburger, et al.**	1994	21.3	18	32		20	0.62	-0.22	1.46
Hatch**	2006	21.1	48	22	4.0	8	0.24	-0.28	0.75
Horak**	2003	11.0	51	54		4	0.10	-0.28	0.49
Hughes*	1994		30	30		1	1.20	0.65	1.75
Jacobson*	1993		16	4	23.5	6	0.72	-0.63	2.07
Jelalian, et al.*	2006	14.5	37	39	16.0	11	0.26	-0.27	0.79
Kanters et al.**	2002		57	78		12	0.25	-0.09	0.59
McDaniel*/**	1999	16.5	22	20	6.0	6	0.37	-0.24	0.98
McDonald & Howe*	1989		18	20	28.0	6	0.52	-0.13	1.17
McGarvey*/**	2004	13.0	29	15		44	-0.04	-0.70	0.62
Miller**	1998		59	68	3.0	2	0.26	-0.09	0.61
Moreau de la Meuse**	1993	12.0	78	63	20.0	5	0.04	-0.31	0.39
Parker*	1993	15.0	37	47	16.0	20	0.10	-0.68	0.88
Shoemaker**	1997	22.2	26	11		1	0.00	-0.71	0.71
Smith**	1994	21.2	20	21		9	0.31	-0.31	0.93
Sturdivant**	1991		14	12	8.0	20	0.17	-0.61	0.96
Sung**	2005		56	56	8.0	2	0.15	-0.38	0.68
Talbot**	2001		30	26		2	-0.10	-0.62	0.43
Thompson-Grim*/**	1999	15.6	17	17	48.2	2	0.18	-0.57	0.93
Ulrich**	1993		7	13	16.0	3	0.31	-0.61	1.24
Vasquez*	2001	14.5	10	11	25.0	2	1.35	0.41	2.30
Voight*	1988	15.0	11	10	5.0	8	0.54	-0.34	1.41
White**	1998	16.1	12	12	45.0	4	0.62	-0.21	1.45
Wisnyai**	1989		27	34		4	0.82	0.08	1.57
Witman*	1987	17.0	5	5	5.0	1	1.04	-0.28	2.37
Witter**	2005	20.2	16	11	3.0	3	-0.24	-1.05	0.58
Ziven*	1989	15.3	36	48		9	0.09	-0.38	0.56

Table 2

Mean effect size, number of effect sizes, percentage of total effect sizes, and confidence intervals for each outcome of therapeutic studies.

Outcome	n (Studies)	%	n (d)	%	Mean d	-95CI	+95CI
Self esteem or Self concept	8	32.0	49	36.8	0.40	0.19	0.61
Self-Efficacy	2	8.0	2	1.5	1.21	0.73	1.68
Personality Measures	5	20.0	20	15.0	0.19	-0.10	0.49
Behavioral Observations	3	12.0	32	24.1	0.39	0.04	0.75
Academic Measures	1	4.0	1	0.8	1.47	0.51	2.44
Family	2	8.0	10	7.5	0.88	0.38	1.37
Physical Variable (e.g., weight, BMI, blood)	1	4.0	4	3.0	0.00	-0.53	0.53
Group Dynamics (Interpersonal, Cohesion, Effectiveness)	3	12.0	15	11.3	0.10	-0.27	0.47

Table 3.

Mean effect size, number of effect sizes and percentage of total effect sizes for each outcome construct of school-based studies.

Outcome	<i>n</i> (studies)	%	<i>n</i> (<i>d</i>)	%	Mean <i>d</i>	-95CI	+ 95CI
Self esteem or Self concept	12	48.0	91	39.4	0.17	-0.09	0.44
Self-Efficacy	5	20.0	32	13.9	0.16	-0.30	0.62
Personality Measures	5	20.0	26	11.3	0.29	-0.13	0.72
Behavioral Observations	3	12.0	30	13.0	0.41	-0.13	0.96
Academic Measures	2	8.0	7	3.0	-0.01	-0.58	0.57
Environment (Atmosphere)	2	8.0	10	4.3	0.10	-0.46	0.66
Attitudes about Physical Condition	1	4.0	7	3.0	0.29	-0.63	1.22
Group Dynamics (Interpersonal, Cohesion, Effectiveness)	5	20.0	28	12.1	0.66	0.27	1.06

Balancing Evaluation and Program Needs

Presented at the Research and Evaluation of Adventure Programming 2009 Symposium: Atlanta, GA

Maurie Lung, LMHC, LMFT
Eckerd Youth Alternatives, Inc./Life Adventures Counseling & Consulting
PO Box 10907 , St. Petersburg, FL 33733
727-481-4577
dmlung@lifeadventurescc.org

Greg Holeman, Chief Quality Officer
Eckerd Youth Alternatives
100 N. Starcrest Dr., Clearwater, FL 33765
727-461-1236 X437
gholeman@eckerd.org

Introduction

The purpose of this presentation was to engage participants in developing and monitoring meaningful organizational outcomes through cross-functional teams and integrating evaluation with practice. An overview of Eckerd Youth Alternatives, Inc. (EYA) process provided a foundation to discuss the potential impact of program outcome evaluation within an organization to encourage responsibility, accountability, and application of evidence-based strategies within the organization. While based on EYA's "real life" learning situation, the focus was on the transfer of strategies to a variety of settings. Strategies discussed included a defined Performance and Quality Improvement Division, developing cross-functional teams, utilizing a Balanced Scorecard™ with key performance indicators, developing a master quality plan and toolkit, and coordinating electronic records and for data collection and reporting functions.

Relevancy to Research and Evaluation of Adventure Programming

Considering the current political and economic climate programs are facing, as well as the desire for programs to continuously strive to provide effective and efficient programs, a program evaluation system is imperative.

Presentation Overview

Eckerd Youth Alternatives, Inc. Organizational Overview

Eckerd Youth Alternatives (EYA) has provided experiential education through a broad base of services (prevention programs, after-school programs, community based programs, residential treatment programs, aftercare programs) to nearly 100,000 youth since its founding in 1968. Currently, there are 38 programs in nine states. EYA's vision is to "ensure that each child has the opportunity to succeed." To do this, EYA's mission is "to develop and share programs that promote the well-being of children and serve at-risk youth and their families." Above all, EYA values the uniqueness and inherent worth of each individual and assisting youth to develop their potential. Related to evaluation and research, EYA has promoted values such as "We embrace innovation, flexibility, reflection, and continuous improvement." And "We hold ourselves accountable for achieving superior outcomes." EYA provides a continuum of care, from prevention/diversion programs, to day treatment programs, to residential programs, to transition/re-entry aftercare programs. These programs may include services that are therapeutic, educational, recreational, or case management related.

Performance and Quality Improvement Division and Charges

The EYA Executive Leadership Team and Board of Directors identified the measurement of performance outcomes, as well as improving the quality of its organizational performance, as key focus areas. A Quality and Performance

Improvement Board Committee and a cohesive Performance & Quality Improvement (PQI) Division (Data Integrity, Information Services, Performance Improvement, Policies & Procedures, and Quality Improvement) were formed.

EYA Leadership team also identified quality, staff, financial, external relations, and growth as the primary spokes of the wheel for the organization, with quality being the hub. Each area had an identified focus. For quality, it is safety, quality assurance, performance improvement and planning, evidence-based practices, and outcomes. For staff it is targeted recruitment, professional development, and retention. For financial it is fiscal strength, discipline, accurate analysis and forecasting, capital allocation, and cash management. For external relations it is donor cultivation, communications, publications, and alumni relations. For growth and transformation it is mission-driven reinvestment, business viability assessment, marketplace analysis, targeted retooling, and expansion.

As a performance-based organization, we clearly define, collect, validate, analyze and communicate our outcomes to make informed decisions based on evidence. Our initial tasks for the PQI division were clear: develop an organizational Balanced Scorecard™ of key performance indicators, develop a Master Quality Improvement Plan and Toolkit, and improve electronic data collection and report functionality.

Cross-Functional Teams

To accomplish our tasks, we created diverse cross-functional teams that included employees from various departments and programs. Although we could have been more efficient by creating these documents ourselves, this created opportunity to increase buy from our employees, increase the accuracy of the key performance indicators, and allowed us to address barriers before we started data collection. This approach was a critical component to success. If you do not have support at the top level of leadership, it may not work. Having a tangible end product and timeline assisted in the process of selling this approach to leadership. Above all else, involve the key decision makers.

To create a formal process, people were selected and invited to attend by a personal contact, including expectations for participation and potential contributions. A kick off meeting was facilitated by EYA leadership to establish the frame for this process. Each team had a facilitator and meetings were held weekly for one hour, with electronic SharePoint space established for each teams work environment and to monitor the tasks. After the process was completed, there was a closing meeting that included a progress update. An interesting note is that after the organization saw this strategy being used well for a way of work, it began to be replicated in other arenas.

Balanced Scorecard™ Cross-Functional Team

The Balanced Scorecard™ was developed by Robert Kaplan and David Norton based on research in 1990 exploring performance measurement in companies. It is a performance management tool to help organizations measure whether the operational activities are aligned with the organization's vision and purpose. Historically, organizations focused only on financial outcomes. However, this concept includes operational, marketing, and developmental inputs to create a more comprehensive view. There are four basic perspectives to identify indicators for: customer/client, internal process, employee learning/growth, and financial. These are further defined as "the customer perspective includes knowing the customer, their expectations and demands from the organization, and the value proposition in serving them. The internal process perspective is what must be excelled at in order to continue to add value for the customers and what processes are in place to best execute the strategies. Employee learning/growth is the level/content of employee skills needed to meet the mission/value, the systems needed to support the processes, and the organizational climate (culture) to support success. The financial perspective is what would help the organization understand the effectiveness and efficiency of the financial processes." (Niven, 2008)

For EYA, the balance scorecard is a tool that ensures that we are consistent in the methodologies used to measure outcomes and are able to provide feedback to all levels of the organization. The primary tasks of this team were to prepare a list of the different scorecards (e.g., Board level, organization level, program level), select the measurement criteria (5-7) per area, identify the source of the data for each measurement, and create the scorecard communication tool. Indicators for performance measures were reviewed by the team to ensure that they were linking cause and effect, understandable, measurable, and driving the "right" behavior. Through this process, this was expanded to create an evaluation plan for each indicator, this included definitions, standard of success (target), instrument, respondents, data collection methods, data verification, analysis, and reporting. In addition to demographics and contract, licensing, grant, accreditation information requirements, for youth outcomes organizationally, EYA uses the Youth Self Report (for mental

health functioning), Matson's Evaluation of Social Skills for Youngsters, Woodcock Johnson-III (for reading and math skills), client satisfaction surveys, and the Eckerd Youth Well-being Instrument to gather information after services.

Master Quality Improvement Plan Cross-Functional Team

The Master Quality Improvement Plan is a comprehensive document that establishes organizational strategies for identifying critical areas, assessing, developing and implementing quality assurance objectives and procedures in a consistent and methodical way. The Plan allows EYA to deliver top quality care to our clients, provide excellent, reliable data and information to our internal and external customers and measures outcomes for continuous quality improvement. The primary tasks of this team were to assess current systems and tools, develop a cross functional needs assessment, define requirements, resources and deliverables, and determine implementation strategies that include training and ongoing support.

Master Quality Improvement Plan Cross-Functional Team

EYA has several key tools that will be used to collect and report evaluation data; this includes the PBViews reporting software and the Totally Integrated Electronic Reporting computer systems. Having the appropriate tools to capture and report large amounts of data is critical in the monitoring of a diverse group of programs in multiple locations. Having effective tools will minimize the time required by staff and leadership to report and analyze quality data. The primary tasks of this team was to deploy the report software with clear expectations on who and when data will be entered, engage the vendor to re-deploy the TIER system, and create a training plan for all users of the different systems to ensure they are comfortable using the system.

Integrating Processes

One example of how EYA has integrated evaluation and operational processes is the Internal Program Review (IPR). The IPR ensures the quality of services for our clients and helps a program to know itself better, to evaluate its own operations more systematically, and to make improvements. EYA's internal program review is a mandatory process of peer evaluation that assesses a program's ability to comply with current industry standards, recognizes a program's strengths or program assets, and provides guidance to a program regarding how to improve practice.

The scope of the Internal Program Review will vary based on the needs of the individual program. The EYA Department of Quality Improvement customizes each program review based on the specific program's licensing requirements, and contract requirements, as well as on EYA policies and COA accreditation standards. The review process may include areas such as, but not exclusively, safety and risk management, therapeutic environment, educational program, programmatic activities, human relations (on-site training, supervision, retention), and additional organizational systems (external reviews and accreditations, program-level quality improvement). A program must meet all applicable standards for which they are currently held accountable (e.g., DJJ, Medicaid, SACS, COA, etc.) Evidence of compliance clearly indicates how each standard is met or why the standard does not apply. The process includes assessment prior to the site visit, a 2-5 day on-site review with a diverse team lead by a Quality Improvement Specialist, and follow-up support that includes connecting with needed identified resources. Following the review, a Quality Improvement Specialist completes a report of findings, both strengths and areas needing improvement. The program's Operations Director and Director develop and implement a corrective action plan. If needed, follow-up site visits are made to determine compliance with unmet standards.

References

Niven, P. (2008). *Balanced Scorecard: step-by-step for government and nonprofit agencies*. Wiley: Hoboken: NJ.

Evaluation and Collaboration: Examining Outward Bound's Research Initiatives

Presented at the Research and Evaluation of Adventure Programming 2009 Symposium: Atlanta, GA

Jon Frankel

Research & Development Manager
Outward Bound Training Institute, Outward Bound USA
910 Jackson Street
Golden, CO 80401
P: 720.497.2356
F: 720.497.2441
jfrankel@outwardbound.org

Alan Ewert, PhD

Department of Recreation, Parks and Leisure Studies
Indiana University
HPER 133
1025 E. Seventh Street
Bloomington, IN 47405-7109
aewert@indiana.edu

INTRODUCTION

Beginning with Shore's (1977) compilation of the effects of Outward Bound (OB), numerous studies and other research efforts have focused on identifying various outcomes that students experience following participation in Outward Bound. The preponderance of these studies has pointed to a positive and growth-enhancing experience. Moreover, numerous variables have been studied including self-concept, trust, recidivism rates, attitude toward school, communication and self-esteem. Few studies, however, have linked the underlying goals of Outward Bound with the actual course outcomes.

This document provides an initial overview of a study which attempts to ascertain the effects of participation on three of the major foundational outcomes of Outward Bound programs: Character Development, Leadership and Service. A second purpose of this study is to test the various properties associated with the instrument developed for, and used in, this analysis.

METHODOLOGY

During summer 2008, questionnaires from 651 students participating in 50 different courses were collected. Data were collected immediately before the course (pre-scores) and immediately following the course (post-scores). There were 633 pre-scores and 630 post-scores collected. Of the total sample, 39 people had either a missing pre or post-test score and were eliminated from further analysis.

Upon inspection of the data, 35 subjects whose Z scores on each variable were higher or lower than ± 3.3 standard deviations (approximately 95% of the distribution of scores) from the mean were deleted from further analysis. Following these actions, 577 (88.6% of the original 651) valid subjects were available for further analysis.

RESULTS

The major content validation methods used to develop this instrument were a review of the literature and the use of two separate expert panels (the Outward Bound Research Advisory Committee and two individual scale designers). Using these approaches suggests an adequate level of content validity.

In this study, the subscales of Character Development, Leadership and Service were all significant and positively correlated with one another, thus suggesting that these subscales are associated with one another. Unfortunately, no variables were measured that would be expected to have a negative correlation with Character Development, Leadership or Service. Therefore, it is unclear whether this dimension of construct validation has been satisfied. Finally, the three variables are so highly correlated that it is unclear whether each one is measuring a distinct psychological trait, thus making any conclusions about content validity problematic.

This survey instrument measures changes in attitude; although ultimately, the criterion of interest may be actual behavior rather than attitude. Due to the lack of longitudinal data, we have no evidence to support the presence of criterion validity at this time.

Reliability

The Cronbach alpha values for the three dimensions (Character Development, Leadership, and Service) fall into the low but acceptable range, thus providing some confidence that the instrument is measuring the three dimensions in a consistent manner. Table A in the Appendix displays the Cronbach alpha values of the subscales and the individual items within each subscale.

Variability

The data from this study present a challenge due to the fact that most responses were grouped at the high-end of the Likert scale—both before and after the course—with every question tending to be positive. Moreover, 85% of all responses were registered as either a 5, 6 or 7 for the pre-test scores and 92% were 5, 6 or 7 for the post-test scores. From the perspective of the mean responses per item, 79% of the mean responses per item on the pretest were 5.5 or higher and 96% of the responses were 5.5 or higher on the post-test scores. While this is perhaps how the respondents honestly felt about their Outward Bound experience, this systematic response also suggests that the outcomes from the instrument are high, in part, because they started high.

Instrument Design Summary

The challenge of any self-report instrument is to accurately reflect the feelings and attitudes of the respondent. Given the challenges of interpretation of the questions, systematic bias and the difficulty in turning a feeling or attitude into a number, the results from any instrument need to be viewed cautiously. This instrument attempts to capture three of the most important foundations of Outward Bound: Character Development, Leadership and Service. Each one of these subscales is composed of a number of individual items, which when taken as a whole attempted to “capture” that particular dimension. While at this stage in the instrument development we do not know about issues related to item interpretation, it would appear that the questionnaire presents acceptable levels of reliability at the subscale level and measures what it purports to measure (validity); however, the lack of variability in the data—due to the ceiling effect—makes the results suspect. As defined by Vogt (1993), the ceiling effect occurs when many of the subjects in a study have scores at or near the possible upper limit (ceiling). The ceiling effect makes analysis problematic because it reduces the amount of variation in the variable being studied, thus inhibiting the finding of any differences between the different scores.

DATA ANALYSIS AND FINDINGS

Mean values were collected for the three subscales and all the individual items. Repeated measures *t* tests were conducted to test the program effects by comparing pre-test and post-test scores. The following results show the result for the overall effects, effects on the individual level variables and effects on the course level variables.

Overall Effects

Paired *t* tests were conducted to see if there were significant differences between pre-test and post-test composite scores; both for the subscales and items. The results indicate significant changes over time on the total score ($t = 20.3$, $p < .01$, $ES = .85$). The results of the paired *t* test for each subscale and items are presented in Table B (see Appendix).

Although the paired *t* tests resulted in significant statistical differences, there are several reasons why the interpretation of the paired *t* tests must be viewed with caution. First, unlike the *t* tests, both multi-level modeling and ANOVA was also used on the data and showed no significant differences between the responses before and after the OB courses. Second, the responses exhibited substantial ceiling effects. As previously mentioned, most responses were at the high end of the 7-point Likert scale, both before and after the courses. Therefore, at this point, any statement that there is a difference between respondent scores before and after courses should be viewed cautiously.

Effects on Individual Level Variables

Two-way repeated measure ANOVA was conducted to see if there was an interaction between change over time and individual level variables such as gender, age and ethnicity. No interaction effects between these groups of individuals were found.

As listed in Table 1, there is significant difference between male and female ($F = 15.39$, $p < .01$). That is, female students reported both higher pretest and posttest scores than did their male counterparts

TABLE 1 – ANOVA Table for Change Over Time and Gender1

	SS	df	MS	F
Change	49.860	1	49.860	392.061**
Change*Gender	.000	1	.000	.003
Residual	73.125	575	.127	
Gender	7.629	1	7.629	15.392**
Residual	284.985	575	.496	

* $p < .05$; ** $p < .01$

In examining age, there is no significant difference between the various groups ($F = .252$, $p = .778$). See Table 2.

TABLE 2 – ANOVA Table for Change Over Time and Age SS

	SS	df	MS	F
Change	36.353	1	36.353	286.327**
Change*Age	.334	2	.167	1.316
Residual	72.243	569	.127	
Age	.256	2	.128	.252
Residual	288.916	569	.508	

* $p < .05$; ** $p < .01$

Using ANOVA, results demonstrated differences among all the ethnic groups combined ($F = 3.288$, $p < .01$; See Table 3). Using a post-hoc LSD test, significant differences can be found between African Americans and Whites (mean difference = $.17$, $p < .05$), and Hispanic or Latino and White (mean difference = $.13$, $p < .05$). There is also a significant interaction effect between change over time and ethnicity ($F = 2.39$, $p < .05$). Upon breaking down the interaction relationship, the results show that in the pretest scores, there are significant differences between White and African American (mean difference = $-.26$, $p < .05$), White and Hispanic or Latino (mean difference = $-.01$, $p < .05$), Asian and African Americans (mean difference = $-.37$, $p < .05$), Asian and Hispanic or Latino, and African American (mean difference = $-.31$, $p < .05$). There are no significant differences in post test among different ethnicity groups. These results suggest that White and Asian students reported higher rates of change over time than African American and Hispanic or Latino students.

TABLE 3 – ANOVA Table for Change Over Time and Ethnicity

	SS	df	MS	F
Change	7.364	1	7.364	26.002**
Change*Ethnicity	4.061	6	.677	2.390*
Residual	161.439	570	.283	
Ethnicity	15.316	6	2.553	3.288**
Residual	442.587	570	.776	

* $p < .05$; ** $p < .01$

However, because White students are the majority group of the sample (65%), Non-white groups were combined and compared to White. The result shows there are no significant differences between White and Non-white on overall changes. However, Non-white students tended to report higher scores on pretest.

Effects on Course Level Variables

Course length, base camp location, activity type and type of course were analyzed using ANOVA. There were no significant differences on the three course length groups (≤ 10 days, 11-16 days and ≥ 21 days).

On base camp location, the ANOVA results (Table 4) show that there are significant differences among base camp locations. The result of LSD post hoc tests showed that there are several pairwise comparisons with significant differences.

TABLE 4 – ANOVA Table for Change Over Time and Base Camp

	SS	df	MS	F
Change	29.864	1	29.864	242.174**
Change*Base Camp	4.191	17	.247	1.999**
Residual	68.934	559	.123	
Base Camp	22.287	17	1.311	2.711**
Residual	270.326	559	.484	

* $p < .05$; ** $p < .01$

For Activity types, ANOVA showed that there are significant differences among different types of activities (see Table 5).

TABLE 5 – ANOVA Table for Change Over Time and Activity Types

	SS	df	MS	F
Change	32.761	1	32.761	264.984**
Change*Activity	5.127	26	.197	1.595*
Residual	67.998	550	.124	
Activity	28.636	26	1.101	2.295**
Residual	270.326	559	.484	

* $p < .05$; ** $p < .01$

SUMMARY AND CONCLUSIONS

Currently, there is some evidence for content and construct validity of this instrument. Conversely, the data lack desirable levels of variability due to a ceiling effect in the responses. A related issue is that all of the statistical tests used in this study assume that the data are normally distributed. In this study, however, the data are negatively skewed. Because of the skewed nature of the data, a non-parametric test to compare pretest and posttest scores (Wilcoxon Signed Ranks test) was also used and compared with the more traditional t test. The results were similar. However, due to the robustness of the t -test, the paired t test was still considered to be the better and easier method for data analysis.

Phase II of this study (taking place in 2009) will address the reliability and variability issues discussed previously. The ceiling effect will be addressed by considering using extreme responses, increasing variability of response (e.g., 1-5, 1-10), data transformation, regression discontinuity designs, and retrospective self-reporting.

Overall, the results point to statistically significant increases in student responses on the subscale dimensions of Character Development, Leadership and Service; the ceiling effect and reliability issues notwithstanding.

References

- Austin, P. C., & Brunner, L. J. (2003). Type I error inflation in the presence of a ceiling effect. *The American Statistician*, 57(2), 97-104.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071.
- Bowen, G. L., Rose, R. A., & Ware, W. B. (2006). The reliability and validity of the school success profile learning organization measure. *Evaluation and Program Planning*, 29(1), 97-104.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Earlbaum.
- Devellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed. Vol. 26). Thousand Oaks, CA: Sage.
- Shore, A. (1977). *Outward Bound: A reference volume*. Greenwich, CT: Outward Bound.
- Vogt, W. P. (2005). *Dictionary of statistics & methodology: A nontechnical guide for the Social Sciences*. Thousand Oaks, CA: Sage.

APPENDIX

TABLE A – Cronbach’s Alpha by Subscale and Individual Item of OBOI

	Pre	Post		Pre	Post		Pre	Post
Character Development	0.67	0.73	Leadership	0.73	0.73	Service	0.69	0.78
SCI	0.54	0.6	GS1	0.52	0.55	SR1	0.62	0.69
SC2	0.62	0.63	GS2	0.61	0.63	SR2	0.65	0.7
SA1	0.54	0.62	GC1	0.52	0.53	ER1	0.57	0.69
SA2	0.63	0.63	GC2	0.6	0.66	ER2	0.69	0.7
CO1	0.49	0.61	EC1	0.61	0.59	AE1	0.59	0.65
CO2	0.51	0.61	EC2	0.55	0.62	AE2	0.64	0.74
HL1	0.58	0.57	CR1	0.56	0.63			
HL2	0.49	0.54	CR2	0.44	0.45			
			PS1	0.59	0.61			
			PS2	0.55	0.6			

TABLE B – T test and Effect Size for Subscales and Factors Subscales

Subscales	Factors	T Score	Effect Size
Character Development		18.4**	0.72
	Self-Confidence	14.9**	0.62
	Self-Actualization	13.8**	0.58
	Compassion	10.0**	0.42
	Healthy & Balanced Life	10.9**	0.46
Leadership		14.7**	0.61
	Goal Setting	15.4**	0.64
	Group Collaboration	8.6**	0.36
	Effective Communication	10.4**	0.43
	Conflict Resolution	2.7**	0.11
	Problem Solving	12.8**	0.53
Service		18.6**	0.77
	Social Responsibility	14.8**	0.62
	Environmental Responsibility	11.0**	0.46
	Active Engagement	14.9**	0.62
Total		20.3**	0.85

* $p < .05$; ** $p < .01$

Evidence-Based Research: Practice and Learnings from Other Fields

Presented at the Research and Evaluation of Adventure Programming 2009 Symposium: Atlanta, GA

Alan Ewert, PhD

Department of Recreation, Parks and Leisure Studies

Indiana University

HPER 133

1025 E. Seventh Street

Bloomington, IN 47405-7109

awert@indiana.edu

While relatively new to the Experiential Education field, Evidence-Based Research (EBR) has a long history in other areas of research and practice. The treatise underlying this presentation is one of caution and reflection. There are issues and concerns as well as positive aspects associated with EBR, that have emerged from its use, that the Experiential Education community might be well-advised to consider as we rush down the Evidence-Based path. It would be prudent to understand some of the history and perspective of EBR in other fields before we move from a respectful handshake to a full-blown lovers' embrace.

For this presentation, I am defining Evidence-Based Research as a systematic evaluation of existing research in a given area for past literature, and rigorous examples of validity, reliability, methodology, design, and data analysis (Chiappelli & Cajulis, 2007). As such, Evidence-Based Research can be considered a process designed to determine the best available method/s or practice/s for achieving program or individual goals through an Experiential Education experience.

Far from being a cloistered academic concept, our contemporary society uses EBR, continuously and often without knowing it. We rely on traditional sources of information for verification of our actions and beliefs, be they from authorities, leaders, or teachers, to newspapers, newsmagazines, and other media. Increasingly, segments of our populations rely on Google, YouTube, and other web-based sources, not just for entertainment but also for information that supports their beliefs and subsequent behaviors.

While less accessible to the public, those of us in academia rely on such esoteric concepts as theory, outcomes based research, validity, and peer-review to demonstrate aspects of evidence-based research and decision-making. But even these stalwart paragons of research efficacy are not without their critics and questions. For example, as we know from the writings of scholars such as Baldwin, Persing, & Magnuson (2004) and Babbie (1998), theories provide the foundational basis of understanding and knowledge. Indeed, Henderson (2004) suggests that "data that lack a theoretical connection might be likened to having a ropes course without a leader" (p. 185). I would differ only in degree, for we all know that ropes course leaders come in all sorts of shapes and sizes and often facilitate the ropes course experience in remarkably different ways. Harkening back to the theory connection, just as it depends on what type and quality of facilitator you get on a ropes course, the way you look at and interpret data often depends on the (type and quality of) theory you happen to be using; and it is this relativism that becomes troublesome. In practice, it is not very helpful to answer the question of "what do the data mean" with "it depends on what theory you are using."

What about the issue of "peer review?" Even the sanctity of this concept is not without its critics. Indeed, no less an august body than the U.S. Supreme Court has recently been wrestling with the issues of acceptability and reliability of scientific evidence and have found the peer review process to be questionable (Horrobin, 2001). Moreover, a study by Rothwell and Martyn (2000) found that agreement between reviewers was not much greater than that expected by chance alone. As a former editor of the *Journal of Experiential Education*, I can attest to the problem of the "split review." But how can reviews be split? Good science is good science and should be plainly evident for all to see... if they know what they're looking for.

Perhaps Saint Basil had something when he stated in his *Hexaemeron*, "At all events let us prefer the simplicity of faith to the demonstrations of reason." But, I personally like Saint Augustine's admonishment and something our graduate students should keep in mind, "The good Christian should beware of mathematicians and all those who make empty prophecies. The danger already exists that the mathematicians have made a covenant with the devil to darken the spirit and to confine man to the bonds of Hell." Well, so much for the comfort of a statistically significant difference to provide some much needed evidence regarding an outcome. Comfort comes only from knowing that quantitative statistics are only one sort of evidence.

However, before we hit too many potholes as we go down the road to evidence-based research let's take a spin on the much well-traveled highway of its origin and virtues. We know for example, that there are numerous professional and scientific endeavors that are "evidence-based." A sampling of these would include the following:

- Evidence-Based Research
- Evidence-Based Inquiry
- Evidence-Based Management
- Evidence-Based Design
- Evidence-Based Decision-Making
- Evidence-Based Policy
- Evidence-Based Guidelines
- Evidence-Based Practice
- Evidence-Based Medicine

In the case of Evidence-Based Medicine there is some evidence that ancient Greece or China were the places of its origin (Sackett, Rosenberg, & Gray, 1996). Although the testing of medical interventions for efficacy has existed since Avicenna's *The Canon of Medicine* in the 11th century, it was not until Professor Archie Cochrane's book *Effectiveness and Efficiency: Random Reflections on Health Services* (1972) that the term "evidence-based" gained traction.

Moving back to the concept of evidence-based research, the National Research Council (2002) has developed some principles for guiding research. These guidelines include the following:

- Pose significant questions that can be empirically investigated.
- Research should be linked to relevant theory and/or conceptual framework.
- Use methods that allow for direct investigation of the research question.
- Provide a coherent and explicit chain of reasoning.
- Be able to replicate and generalize across studies
- Disclose research for professional scrutiny and critique.

To be sure, the National Research Council pointed out a wide variety of legitimate scientific designs ranging from experiments to in-depth qualitative case studies (McMillan & Schumacher, 2006, p. 9). But to be *scientific*, the design must allow empirical (i.e., evidence-based) investigation of the question under study.

As these guidelines suggest and most of us already know, the evidence-based research system is definitively tilted toward an empirical, positivistic, perspective with the "gold standard" increasingly being the randomized controlled trials, with relatively large numbers of subjects and with the sources of bias and confounding variables controlled for (Ewert & Sibthorp, 2009). In addition, the concepts of treatment fidelity or methodological strategies for monitoring and enhancing the reliability and validity of behavioral interventions (Borrelli, et al., 2005), and carefully defined interventions are also playing more important roles in the development of evidence-based research.

Within this paradigm, there exists different levels of qualifications of evidence. For example, from an intervention perspective, the "strongest" evidence for therapeutic efficacy is provided by a systematic review of randomized with control, double-blind, placebo-controlled involving a homogeneous population with the same set of issues or conditions. Student testimonials, case reports, and expert opinion have much less value because of issues such as the placebo effect, inherent biases, difficulties in discerning who is an expert, social desirability, and so forth.

Extending on this line of reasoning, consider the system set up by the U.S. Preventive Services Task Force for ranking evidence regarding the effectiveness of treatments.

- Level I: Evidence obtained from properly designed randomized controlled trial.
- Level II-1: Evidence from a well-designed controlled trial without randomization.
- Level II-2: Evidence obtained from a well-designed cohort or case-controlled analytic study from more than one research group.
- Level II-3: Evidence obtained from multiple time series. Dramatic results may also be considered as this type of evidence.
- Level III: Opinions of respected authorities, descriptive studies, or expert committees.

Our colleagues in the UK have developed a similar system, different in terms, but almost identical in intent (Atkins, Best, Briss, et al., 2004).

- Level A: Consistant randomized controlled studies. Decision rules validated in different populations
- Level B: Consistent Retrospective Cohort, ecological study, outcomes research, case-control studies
- Case-series studies
- Expert opinion without explicit critical appraisal

Finally, what is the connection between Evidence-Based Research and Experiential Education as we understand it? Are there some functions, either managerial or organizational, in which Evidence-Based Research can be particularly effective? While not definitive, consider that following possibilities:

- How are managerial and pedagogical decisions made in EE situations?
- In what ways can Evidence-Based Research actually contribute to organizational performance and course outcomes?
- Are there specific organizations or structures in EE that are particularly facilitative in the acquisition or use of Evidence-Based Research?
- What is the likelihood that Evidence-Based Research will actually be used in EE organizations and programs?

The Problem

As can be seen from the aforementioned section, Evidence-Based Research can provide a powerful decision-making tool in ascertaining the effectiveness and efficacy of a particular program design or treatment modality. Not surprisingly, however, this strength also becomes one of the criticisms leveled at it. In this section, I will attempt to list some of the major concerns and criticisms leveled at the process of Evidence-Based Research.

As previously noted, Evidence-Based Research attempts to provide knowledge that is rigorous, systematic, and objective (Redfield, 2004). As a result, research findings that fall into this rubric are often selective and specialized and as a result, somewhat exclusive. That is, not all the domains of Experiential Education can be handled in an Evidence-Based paradigm. For example, in the Experiential Education setting it is often not possible, ethical, or even desirable to utilize a traditional randomized and well-controlled set of trials or experiments.

This does not mean or even imply that Experiential Education is an incomplete field, rather, like many other fields such as education, it is a practice-oriented profession rather than strictly a theoretical one. Evidence-Based Research, however, often applies more to populations than it does to the student standing in front of you. There is a difference between Experiential Education based on evidence and Evidence-Based Experiential Education, with the latter being much more restrictive and prescriptive.

Simply using a "level of evidence" as the determining factor for practice ignores the importance of integrating the practitioners experience with past literature and knowledge as well as the researcher's data (Glick, 2007). One example of this is a recent paper examining the need for purifying water in Wilderness settings (Webb, 2004). Using "Evidence"-Based Research it was concluded that water purification was not needed but that hand-washing was. Numerous practitioners would probably have serious reservations about this policy based on their past experience and history. Perhaps attention to this latter type of evidence would have been useful.

Beyond this, some would question the epistemological assumptions underlying Evidence-Based Research. Some authors suggest that Evidence-Based Research is a "retrograde scientific process" that brushes aside qualitative methodologies and alternative epistemologies in favor of "old process-product paradigms" such as strict experimentation and adherence

to the "gold standard" (St. Pierre, 2006). As such, critics of Evidence-Based Research suggest that knowledge and processes that are deeply embedded in the social, culture, political, and economic milieu of a particular setting are often overlooked or discarded as inconsequential. In addition, Evidence-Based Research does not extinguish the problem of extrapolation to different populations, settings, or methods. Questions always remain as to how does this research apply to my situation, with my students, and with my staff.

Finally, funded and/or published studies using Evidence-Based Research may not always elucidate the most efficacious or "doable" practice. In part, this is true, because funded grants and published articles often tend to replicate the stated belief system rather than exploring new terrain. Perhaps Neal Miller, former president of the American Psychological Association said it best, "*Published reports of research are written with the wisdom of hindsight. They leave out the initial blind groping and fumbling to save journal space. . . Therefore, they present a misleading picture, which is far too orderly and simple, of the actual process of trying to extend the frontiers of science into unknown territory.*" (in Goran, 1974, p. 81).

Not surprisingly, Evidence-Based Research has difficulty capturing the kind of tacit and craft-like knowledge that undergirds much of what we teach and how we teach through Experiential Education (Polanyi, 1998). Lack of evidence and lack of benefits are not the same.

Conclusions

So, what can we say about Evidence-Based Research in Experiential and Adventure Education? Certainly, Evidence-Based Research can and will play an important role in the illumination of outcomes realized through participation in Experiential Education activities and experiences. Perhaps, what Evidence-Based Research really represents is good research with all its attendant issues of sampling, validity, reliability, appropriate analysis, and theory-building but with more attention shown to developing a more collective understanding of what works and what doesn't. As Green and Britten (1998) and Dirx (2006) remind us, fields such as medicine and education are more than just the application of scientific rules. Rather, they also embody applied experience, based on personal observation, reflection, and judgment. And it is the integration of all these factors that will develop more effective Experiential Education experiences. Thus, Evidence-Based Research fits into a broader schemata of developing our understanding of what works, what doesn't, and why in Experiential Education. As such, it is part of a constellation but not the entire sky.

References

- Atkins, D., Best, D., Briss, P. A., et al., (2004). Grading quality of evidence and strength of recommendations. *BMJ* 328, 1490.
- Babbie, E. (1998). *The practice of social research* (8th ed.). Belmont, CA: Wadsworth.
- Baldwin, C., Persing, J., & Magnuson, D. (2004). The role of theory, research, and evaluation in Adventure Education. *Journal of Experiential Education*, 26(3), 167-183.
- Borrelli, B., Sepinwall, D., et al., (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology*, 73(5), 852-860.
- Chiappelli, F., & Cajulis, O. S. (2007). Transitioning toward evidence-based research in the health sciences for the XXI Century. *Evidence-based complementary and alternative medicine*, 5(2), 123-128.
- Dirx, J. M. (2006). Studying the complicated matter of what works: Evidence-based research and the problem of practice. *Adult Education Quarterly*, 56(4), 273-290.
- Ewert, A., & Sibthorp, J. (2009). Creating outcomes through Experiential and Outdoor Education: The challenge of confounding variables. *Journal of Experiential Education*. 31(3), 376-389.
- Glick, M. (2007). The clinical expert: An empiric oddity. *JADA*, 138, 432-434.

- Goran, M. (1974). *Science and Anti-Science*. Ann Arbor, MI: Ann Arbor Science.
- Green, J., & Britten, N. (1998). Qualitative research and evidence based medicine, *BMJ*, 316, 1230-1232.
- Henderson, K. A. (2004). Got research in Experiential Education? Theory and practice. *Journal of Experiential Education*, 26 (3), 184-189.
- Horrobin, D. (2001). Something rotten at the core of science. *Trends in Pharmacological Sciences*, 22(2), 51-52.
- McMillan, J. H., & Schumacher, S. (2006). *Research in education: Evidence-based inquiry* (sixth edition). Boston: Pearson Education.
- Polanyi, M. (1998). *Personal knowledge: Towards a post-critical philosophy*. London: Routledge.
- Redfield, D. (2004). A closer look at scientifically based research: How to evaluate educational research. *T.H.E. Journal*, 31, 24-25.
- Rothwell, P. M. & Martyn, C. N. (2000). Reproducibility of peer review
in clinical neuroscience: Is agreement between reviewers any greater than would be expected by chance alone? *Brain, A Journal of Neurology*, 123(9), 1964-1969.
- Sackett, D. L., Rosenberg, W., Gray, J. A. et al., (1996). Evidence based medicine: What is it and what it isn't. *BMJ*, 312, 71-2.
- St. Pierre, E. A. (2006). Scientifically based research in education: Epistemology and ethics. *Adult Education Quarterly*, 56(4), 239-266.
- Webb, T. R. (2004). Evidence-Based Medicine in the Wilderness: The safety of backcountry water. *Wilderness and Environmental Medicine*, 15, 235-237.