5-11-2023

Developing and establishing validity evidence for a measure of Situational Leadership in an outdoor leadership context

Guy B. DeBrun
James Madison University

Developing and Establishing Validity Evidence for a Measure of Situational Leadership in an Outdoor Leadership Context

Guy B. deBrun

A dissertation submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

Department of Strategic Leadership Studies

May 2023

FACULTY COMMITTEE:

Committee Chair: Ben Selznick

Committee Members/Readers:

Christine DeMars

Karen Ford

Table of Contents

Dedication

This dissertation is dedicated to God, my family, professors, and the outdoor education community who have supported and inspired me.

First and foremast I dedicate this work to God, the giver of wisdom. “For the Lord gives wisdom; from his mouth come knowledge and understanding” (*The Ryrie Study Bible*, 1978, Proverbs 2:6).

To my family, my wife, Karen, sons Colden and Wyatt, and parents Claire and Harlan. To Karen for her unwavering love and support, without which I would not have completed this degree. To Colden and Wyatt for keeping me humble by asking, “what would a strategic leader do?” To my parents, for their nurturing care.

To the following faculty at James Madison University. To my advisor Dr. Ben Selznick, for your instruction, support, enthusiasm, and guidance throughout my program of study and this project. To my committee Dr. Karen Ford and Dr. Christine DeMars for sharing your expertise and giving your time. Also, to everyone from the Center for Assessment and Research Studies who generously made time for me, including Dr. Debbi Bandalos, Dr. Caroline Prendergast, and Dr. Keston Fulcher.

Finally, I dedicate this dissertation to the outdoor educators throughout the world investing daily in the next generation of leaders.

List of Tables

List of Figures

Abstract

Discussions of what it means to be an effective outdoor leader are common in outdoor education literature (Smith & Penny, 2010). Research has identified core competencies, conceptual frameworks, and course curricula for effective leadership. However, the criteria upon which judgements are made about leaders lack clarity (Smith & Penny, 2010). Furthermore, very little is documented, and few evaluation instruments exist to evaluate outdoor leader effectiveness (Phipps et al., 2005). The lack of instruments is problematic for four reasons: (1) it hampers efforts to create recognized outdoor leader certifications, (2) it impedes efforts to demonstrate the value of outdoor leadership program outcomes, (3) it hinders the ability to provide feedback to students, and inform leadership curriculum design, (4) it makes pairing outdoor leadership teams a matter of guesswork.

The Outdoor Situational Leadership Rubric (OSLR) was the focus of this study. The rubric was designed to measure college students' knowledge of four basic leadership styles put forth in the Situational Leadership Model, and the ability to assess the performance readiness of followers. Students participating in the outdoor leadership training were required to keep a structured journal throughout the course. A four-element, behaviorally anchored assessment rubric (the OSLR) was used to evaluate the major components of the Situational Leadership Model using students' structured journals.

Generalizability theory was used to gather information on the relative magnitude of different sources of error the OSLR and produce dependability coefficients. The results of this study do not support the assertion that faculty members can dependably rate situational leadership using the OSLR. It may be that lack of true score variance contributed to the low G and Phi coefficients. More research is needed to examine the

psychometric properties of the OSLR. In addition, these findings suggest that further scrutiny of the conceptual basis of Situational Leadership may be warranted.

Chapter 1

Introduction

According to Priest and Gass (2018), "Outdoor leadership is a process of social influence in an outdoor setting where the outdoor pursuits are the media used to create the adventure experiences" (p. 11). Outdoor leadership programs were born from the tumult of 1940's World War II Europe. It was observed that young sailors were dying at greater rates than older, more experienced sailors at the hands of German U-boats torpedoing merchant marine and naval vessels in the North Atlantic (Minor, 1990, as cited in Priest & Gass, 2018). In response, German educator Kurt Hahn partnered with Lawrence Holt, director of a shipping line, and British educator Jim Hogan to serve the needs of young sailors and youth (Priest & Gass, 2018). Their work focusing on providing youth with an experience to foster self-confidence and improve their ability to work with others became known as Outward Bound (Priest & Gass, 2018).

After the war Hahn's attention turned toward addressing a variety of challenges facing youth of the day including declining fitness, initiative, enterprise, memory, imagination, skill, and care (Richards, 1990). During the 1950's, Outward Bound spread across the United Kingdom. In 1958 Outward Bound opened a school in Malaysia, followed in 1961 by the first in the United States, the Colorado Outward Bound school. Today Outward Bound operates 38 schools in 35 countries spread across 6 continents (Outward Bound International, n.d.).

Recognizing a need to train leaders for Outward Bound, Paul Petzoldt and Ernest "Tap" Tapley created the National Outdoor Leadership School (NOLS) in 1965 (Bachert, 1990). NOLS evolved to focus on teaching leadership to thousands of students all over

the world. In 1976, Petzoldt began leading groups of students from Midwestern universities on outdoor leadership training programs in Wyoming. From these efforts sprang the Wilderness Education Association (WEA) (Lupton, 1990). The WEA's mission is, "to promote the professionalism of outdoor education and leadership by establishing standards through curriculum design, implementation, advocacy, and research-driven initiatives" (Wilderness Education Association, n.d.). Today the WEA is a member organization focused on developing standards, accreditation, and certification of outdoor leaders broadly.

Outdoor leadership emerged in American colleges and universities through outing clubs founded in the late 1800's and early 1900's. The Alpine Club (1863) at Williams College was founded to explore interesting places, become acquainted with natural history, and improve pedestrian powers of the members (Morgan, 1999, as cited in Webb, 2001). In 1915 student members of the Alpine Club created the Williams Outing Club. This club's charter included developing personal initiative and leadership among its aims (Webb, 2001). Pre-dating Williams' club by six years, the Dartmouth Outing Club (1909) listed the development of personal traits such as initiative, integrity, self-reliance, and leadership as important goals in its constitution (Webb, 2001). In 1932 the Intercollegiate Outing Club Association was formed with 14 members. Over the course of history, it is likely that student-created outing clubs existed at three to five hundred colleges and universities in the United States (Webb, 2001). It is noteworthy that leadership development is recognized as an outcome at the earliest inception of these programs. For a variety of reasons, outing clubs declined in the 1960's (Webb, 2001), however the void they left was filled by campus recreation outdoor programs.

Campus recreation outdoor programs are frequently housed within university recreation departments but are also found in student organizations, residential communities, or student leadership programs (Speelman & Wagstaff, 2015). Campus recreation outdoor programs are distinguished from outing clubs in that they typically have greater control and oversight from professional staff housed within university recreation or student affairs. These programs frequently include student-led outdoor trips, indoor climbing walls, challenge courses, and outdoor equipment rental operations. Many programs provide students leadership opportunities with campus recreation outdoor programs including serving on steering committees, leading trips, and planning events (Andre et al., 2017). There are currently 300 known campus recreation outdoor programs (NIRSA, 2021).

Outdoor orientation programs (OOPs) are emerging across postsecondary institutions in the United States, serving over 25,000 students each year at 191 American colleges and universities and are steadily increasing (Bell et al., 2014). OOPs are orientation or pre-orientation experiences for small groups (15 or fewer) of first-year students that use adventure activities and include at least one overnight in a wilderness setting (Bell et al., 2010). Seventy-three percent of OOP program administrators identified enhancing student leadership skills as a goal of their program (Galloway, 2000).

Campus recreation outdoor programs and OOPs are examples of peer leadership programs. The objective of peer leadership programs is to provide services to students participating in the programs. However, a significant benefit occurs for the students leading these programs. For example, peer mentors who serve first-year students report

increased confidence in their ability to manage group dynamics, facilitate learning, and empathize with their students (Harmon, 2006). Orientation leaders report growth in their ability to communicate effectively, lead groups, and work under pressure (Russel & Skinkle, 1990).

Another area of outdoor leadership prevalent in American Postsecondary Education is degree-granting outdoor leadership programs. Names for these programs include adventure education, outdoor experiential education, adventure programming, and outdoor leadership (Seaman et al., 2017). Citing the Society of Park and Recreation Educators, Attarian (2001) noted an increase in academic programs from 17 in 1987 to 41. Using mailing lists from the Association for Experiential Education (AEE) and the Association of Outdoor Recreation and Education (AORE), internet resources, and personal knowledge, Seaman et al. (2017) identified ninety-six programs.

Seaman et al. (2017) surveyed outdoor employers' and faculty of outdoor education programs' preferred qualifications of students. Employers selected leadership skills as most important among the top three most valuable skills for new employees. Faculty identified leadership skills as the most valuable qualifications to employers. Seaman et al.'s (2017) discovery of 96 academic outdoor programs is the highest number found to date. This suggests that outdoor activity popularity remains strong, and programs have gained legitimacy within post-secondary education. Seaman et al.'s (2017) findings further support the importance attributed to leadership training in outdoor programs.

The growth of organizations like Outward Bound, and NOLS indicate an interest in leadership training in an outdoor context worldwide. This trend is paralleled in

American postsecondary education. Campus recreation outdoor programs and outdoor orientation programs, have grown since their introduction by way of outing clubs in the 1900's. Furthermore, outdoor education academic programs are on the rise. A common thread throughout these programs is leadership development, a concept I turn to next.

Researchers dedicated considerable attention in the 1980's to identify the core competencies of outdoor leadership (Buell, 1981; Green, 1981; Priest, 1986; Priest, 1987; Raioloa, 1986; Swiderski, 1981). This research resulted in twelve core competencies (Priest & Gass (1997, 2018) thought to make up the construct of effective outdoor leadership. Priest and Gass' (1997, 2018) competencies have been accepted with some degree of consistency by several authors (see, Graham, 1997; Kosseff, 2003; Martin et al., 2006; Ogilvie, 2005) (Smith & Penny, 2010). Scholars extended Priest and Gass' work by going beyond skill-based competencies to include either original theories (Jordan, 1989; Priest & Dixon, 1991) or incorporation of broader leadership theory (Brymer & Gray, 2006; Smith & Penny, 2010).

Efforts to define the scope of outdoor leadership are an important step; however, without the ability to measure leadership, the field is left at a standstill. Crawford and Kelder (2017) noted that within leadership studies, leadership assessment has been challenging; many practitioners reject instruments conceived by scholars. Instead, they adopt their own, less rigorous scales. For example, 63.5% of US businesses use internally-constructed 360-degree assessments (Crawford & Kelder, 2017). This problem is mirrored in the outdoor leadership field. The absence of rigorously designed measures of leadership leaves most programs to develop their own measures or use existing tools (Bobilya et al., 2017). The existing tools frequently lack validity or reliability evidence.

In addition, a notable limitation of studies seeking to demonstrate leadership skill improvement through participation in outdoor education is the failure to measure actual leadership behaviors. Many studies relied on self-reported leadership skills or perceptions (Ewert & Overholt, 2010; Frauman & Anderson, 2018) rather than actual leadership behaviors. Other studies (Starbuck & Bell, 2017; Huey et al., 2014; Propst & Koesler, 1998) focused on antecedents of leadership such as self-efficacy and confidence. However, belief in one's ability does not always correlate with actual ability (Huey, 2014). This study addressed the need for a rigorous measure of leadership within the outdoor leadership context.

Background of the Problem

Discussions of what it means to be an effective outdoor leader are common in outdoor education literature (Smith & Penny, 2010). Research has identified core competencies, conceptual frameworks, and course curricula for effective leadership. However, the criteria upon which judgements are made about leaders lack clarity (Smith & Penny, 2010). Furthermore, very little is documented, and few evaluation instruments exist to evaluate outdoor leader effectiveness (Phipps et al., 2005). In addition, the perception that leadership is too difficult to measure has impeded efforts to establish certification of outdoor leaders (Attarian, 2001; Priest, 1988; Priest & Gass, 2018).

The lack of instruments is problematic for four reasons: (1) it hampers efforts to create recognized outdoor leader certifications, (2) it impedes efforts to demonstrate the value of outdoor leadership program outcomes, (3) it hinders the ability to provide feedback to students, and inform leadership curriculum design, (4) it makes pairing

outdoor leadership teams a matter of guesswork. Each of these problems will be presented in greater detail in the following section.

The issue of certification of outdoor leaders has been a source of controversy and debate since the 1970's (Gass, 1999). Medical and hard-skill (rock climbing, kayaking, caving, etc.) certifications have become common industry standards. However, a general "blanket" certification process for outdoor leaders has not been recognized; the profession has favored program accreditation instead (Priest & Gass, 2018). Part of the objection to certification stems from the perception that soft or "people skills" like leadership are difficult to adequately identify, train, and evaluate (Attarian, 2001). Priest (1988) went so far as to say, "Let the certificate be one of skills and not one of leadership" (p. 42). A rigorous instrument designed to measure leadership could help pave the way to acceptance of outdoor leader certification.

Scholars and practitioners lack a reliable tool designed to measure leadership outcomes of outdoor programs. One way to communicate the value of outdoor recreation programs is by quantifying the contribution the program makes to the educational mission of the institution. Educating and developing students as leaders is a central purpose for institutions of higher education, as evidenced in mission statements and by the increased presence of both curricular and co-curricular leadership development programs (Astin, 1993). An effective measure of outdoor leadership provides a means to measure leadership outcomes.

Instructors of academic outdoor leadership programs and administrators of outdoor programs lack the means by which they can provide evidence-based feedback to their students and inform curriculum and training design. Outdoor leadership as a

profession has been slow in developing effective ways of assessing student knowledge and performance in outdoor leadership preparation programs (Pelchat & Karp, 2012). Little research has been done on integrating competencies into curricular, instructional, assessment design, and application which has resulted in erratic development of outdoor leadership training programs (Pelchat & Karp, 2017). The instrument proposed for this study measured outdoor leadership competency and therefore could be used in studies designed to address this gap in the literature.

Outdoor leadership requires leading activities which include a high degree of perceived and/or inherent risk in remote places for lengthy time periods (Jordan, 1989). Therefore, it can be an emotionally, mentally, psychologically, and physically demanding form of leadership (Rilling & Jordan 2007), where trip leaders frequently work in pairs. The dynamics of outdoor leadership pairs has a great effect on the outdoor adventure experience (Rilling & Jordan, 2007). Therefore, a tool that provides valid and reliable information on trip leaders' leadership strengths and weaknesses could be leveraged by program administrators in the selection, partnering, and mentoring of outdoor leadership teams.

Research on outdoor leadership has been conducted using existing instruments (Brymer et al. 2010; Ewert & Overholt, 2010; Hayashi and Ewert, 2006). Phipps (1986) developed a measure and method based on Situational Leadership Theory. Each of these instruments is designed to measure leadership constructs and has great utility in understanding outdoor leadership. However, none of these instruments assess leadership as it is practiced by students in the field.

Outdoor leadership is complex, the assessment of which is amenable to performance-based authentic assessment. Authentic assessments are direct assessments of complex performances, tasks valued in their own right (Linn et al., 1991). Examples include hands on problems, essays, and computer simulations. By contrast, multiple-choice tests are indicators or correlates of other valued performances (Linn et al., 1991). The unique contribution of this study is to provide an authentic assessment instrument designed to directly measure leadership as it is being practiced in the field.

The Standards for Educational and Psychological Testing (Standards) (2014) define rubrics as, "the established criteria, including rules, principles, and illustrations, used in scoring constructed responses to individual tasks and clusters of tasks" (p. 223). Per Popham (1997) the critical components of a rubric are: evaluative criteria, quality definitions, and a scoring strategy. Evaluative criteria distinguish the acceptability of responses. Quality definitions explain how to judge qualitative differences. Scoring strategies include aggregating for a single score or scoring by each criterion. This study will develop a rubric which will be used to assess leadership during an experiential leadership experience.

Significance of the Study

This study is significant in several ways. First, many outdoor education programs view leadership development as a principal outcome of the program (Ewert & Overholt, 2010). There is wide support for the importance of leadership in outdoor education. However, little is known about whether leadership skills are actually improved through participation in outdoor education programs (Ewert & Overholt, 2010). A measure of

actual leadership behaviors is needed for scholars to test the efficacy of outdoor leadership programs.

Demonstrating the value of campus outdoor recreation programs is important in today's political climate. The media (Blumenstyk, 2012; Friel, 2003; Martin, 2012; Scott, 2012; Woodhouse, 2015) and public view campus outdoor recreation programs and facilities as examples of excess in higher education (Andre et al., 2017). Directors of campus outdoor recreation programs will need to be able to communicate the value of their programs to school administration, political decision makers, the media, and their constituents (Andre et al., 2017). A measure of outdoor leadership is necessary for practitioners to make claims about contributions to student leadership learning outcomes of campus recreation programs.

Furthermore, research has demonstrated that outdoor education can positively affect participants (Hattie et al., 1997). Povilaitis et al. (2019) note that instructor leadership traits have been shown to influence student outcomes and participants have frequently identified the instructors as important in the learning process (e.g., Asfeldt & Hvenegaard, 2014; Coley et. al., 2015; McKenzie, 2003; Paisley et al., 2008). Leadership traits were not the focus of this study. Yet, if instructor leadership traits are one important factor in participants' achievement of positive outcomes, then a measure of leadership effectiveness is necessary to improve leadership skills.

Research Questions

This study aimed to answer the following research question: To what extent will the newly developed Outdoor Situational Leadership Rubric (OSLR) produce reliable and valid scores of situational leadership amongst participants in outdoor leadership training

programs? The primary research question was, can outdoor education faculty members dependably rate situational leadership using a newly designed rubric? This question was answered using G-theory. G-theory replaces the more well-known concept of reliability. Dependability is how well observed scores allow generalizations across different sets of conditions (Bandalos, 2018). For example, if all of the raters use the OSLR in a similar way, and if each of the items are similar in difficulty, the amount of error attributed to raters and items will be small. It would not be necessary to qualify the interpretation of a score with the specific rater or item. Thus, we will have evidence that scores will generalize across raters and items and are therefore dependable. There are two hypotheses for this study:

Hypothesis 1: The G and phi-coefficients for the scores will be adequately high (> .80)

Hypothesis 2: The largest source of the variance will be subsumed by students, (σ_s^2) which is the object of measurement.

Phi-coefficients are often lower than G-coefficients because they incorporate more sources of error. Thus the .80 cutoff is a rigorous standard.

Purpose of the Study

The purpose of this study was to develop items and gather validity evidence for a new measure of outdoor leadership: the Outdoor Situational Leader Rubric (OSLR). The OSLR is designed to measure knowledge and application of situational leadership. This study proposed to establish reliability and validity evidence for the OSLR.

The following chapters review the literature and describe the methods for this study. The literature review places the current study in the context of outdoor leadership

research, describes current methods for measuring leadership, and identifies Situational Leadership as the theoretical framework. The methodology describes the procedure, sample, and introduces the primary focus of this study, the Outdoor Situational Leadership Rubric.

Chapter 2

Review of Research and Theory

Attempts to establish guidelines for the training of outdoor leaders in the United States date as far back as 1977 (Buell, 1981). Research in this area can be organized into three distinct categories. First, the early studies focused on finding the competencies necessary for effective outdoor leadership and determining the appropriate curriculum for outdoor leadership preparation programs. Next, scholars considered how mainstream leadership theories such as transformational and situational leadership fit into the outdoor leadership landscape. In some cases, authors built on these theories to introduce new outdoor leadership theories. Finally, attention shifted to teaching and measuring leadership as an outcome of outdoor leadership programs.

Researchers dedicated considerable attention in the 1980's to identifying the core competencies of outdoor leadership (Buell, 1981; Green, 1981; Priest, 1986; Raioloa, 1986; Swiderski, 1981). This review will examine the classic research on outdoor leadership competency. Next, additional contributions to outdoor leadership theory (e.g., Comprehensive-Interaction-Expectation model; Jordan, 1989, Conditional Outdoor Leadership Theory; Priest & Dixen, 1991) will be examined. Contingency leadership theories, especially situational leadership, are a common theme in this literature. Situational leadership is used to teach leadership by prominent outdoor leadership organizations (e.g., NOLS, WEA), and is influential in outdoor leadership theory (e.g., Conditional Outdoor Leadership Theory; Priest & Dixen, 1991). An overview of situational leadership and its place in outdoor education follows. Finally, the literature on

teaching and measurement of leadership outcomes within outdoor leadership is examined.

Classic Research

Outdoor leadership research conducted in the 1980's gave considerable attention to determining the qualities of effective outdoor leaders (see Table 1). Priest's (1986) review of this research concluded that identifying important elements of effective outdoor leadership is possible (Priest & Gass, 2018). From this work Priest and Gass (1997, 2005, 2018) presented 12 core competencies. These core competencies have been adopted with some degree of consistency by several authors (Graham, 1997; Kosseff, 2003; Martin et al., 2006; Ogilvie, 2005) and gained traction as a reference point in outdoor leadership training (see Stremba & Bisson, 2009) (Smith & Penny, 2010). The following section reviews the classic research in chronological order.

Buell's (1981) research explored essential, important, and desirable competencies for both entry-level and experienced outdoor leaders. Buell (1981) identified eight competencies deemed essential for entry-level professionals. The study identified an additional 60 competencies deemed essential for experienced professionals. Green (1981) used a Delphi consensus to develop a college course curriculum for land-based outdoor leadership. Green's work identified 35 topics. Green's work helped the field begin to understand what some experts in a specific region of the United States considered important for outdoor leaders. Swiderski (1981) identified 50 important competencies and discovered regional differences in the perceptions of relative importance of competencies. From 39 competency areas Priest (1984) identified nine considered to be of "great importance" and 24 "important." Railoa (1986) sought to establish, test, and

evaluate a curriculum for outdoor leadership. Railoa's research resulted in nine elements considered important for outdoor leaders.

Priest (1986) drew on previous research (Buell, 1981; Green, 1981; Priest, 1984; Swiderski, 1981) and related literature to create a list of 14 outdoor leadership competencies. Priest (1986) developed a survey instrument using the 14 competencies and presented it to outdoor experts representing five different nations (Australia, Canada, Great Britain, New Zealand, and the United States of America). Priest (1986) found that experts from the five nations agreed that safety skills, judgment based on experience, awareness and empathy for others, group management skills, and problem-solving skills were the five most important.

These studies are similar in that they were all dissertations that used experts to identify important skill competencies for outdoor leaders. The lack of peer-review of the early studies must be acknowledged as a weakness. There are several concerns regarding the representativeness of the samples. For example, with the exception of Priest (1984), the samples were heavily skewed towards men. None of the studies reported racial demographics. The apparent lack of diversity of the early studies may contribute to the disconnect between outdoor leadership training and social justice observed by some authors. Warren (2002) explained, "Outdoor leaders who are well trained in communication and group leadership skills, and highly experienced in technical skills, are often at a loss as to how to address social justice issues that arise on courses" (p. 231). These limitations notwithstanding, the classic studies became the foundation on which much outdoor leadership training is based.

Priest and Gass (2018) identified twelve core competencies of effective outdoor leadership from Priest's (1986) analysis that became an evidenced-based curriculum. The twelve competencies are: technical skills, safety skills, environmental skills, organizational skills, instructional skills, facilitation skills, flexible leadership style, communication, professional ethics, decision making, problem solving, and sound judgment. Of particular relevance to this study is flexible leadership style. Priest and Gass (2018) championed the Conditional Outdoor Leadership Theory (COLT; Priest & Chase, 1989). COLT is a contingency theory, incorporating elements of Situational Leadership; this will be explored in greater detail in the following sections.

Priest and Gass' (2018) work has received some criticism. Brown (2004) questioned the efficacy of the leader as a facilitator who assists participants in making their own meaning from experience by guiding discussions, a view espoused by Priest and Gass (1997, 2018) among others. Brown (2004) argued that this advances a false conception of the leader as a neutral background figure and that students are not free to draw meaningful and valid conclusions from their own experiences in leader-guided discussions. Brown (2009) challenged the central idea embraced by Priest and Gass (2018) that learning is highly individual and that lessons gleaned in an adventure experience are portable and may be transferred to "real life." In contrast, Brown (2009) argued that learning is contingent on people, place, activity, and culture. The learner is freed from the expectation to demonstrate learning or new insights that might or might not be applicable in the 'real life" circumstances of home, school, or workplace (Brown, 2009).

Smith and Penny (2010) contended that Priest's and Gass' (2018) approach was too narrowly focused on the outdoor educator, as opposed to the outdoor leader, resulting in a largely skills-based approach. Smith and Penny (2010) believed this perspective compromises the depth of insight into effective outdoor leadership.

As previously mentioned, Priest and Gass' (2018) work was based on studies whose samples lacked diversity. In Warren's (2002) content analysis of outdoor leadership texts, the author concluded that Priest and Gass' (1997) work ignored or marginalized social justice issues. It should be noted that Priest and Gass' third edition (2018) appears to have addressed many of Warren's concerns.

Contemporary theories

Several authors have added to the discussion on effective outdoor leadership since the classic work of the 1980's. These authors distinguish themselves from the early work by going beyond lists of skill-based competencies to include either original theories or incorporation of broader leadership theory.

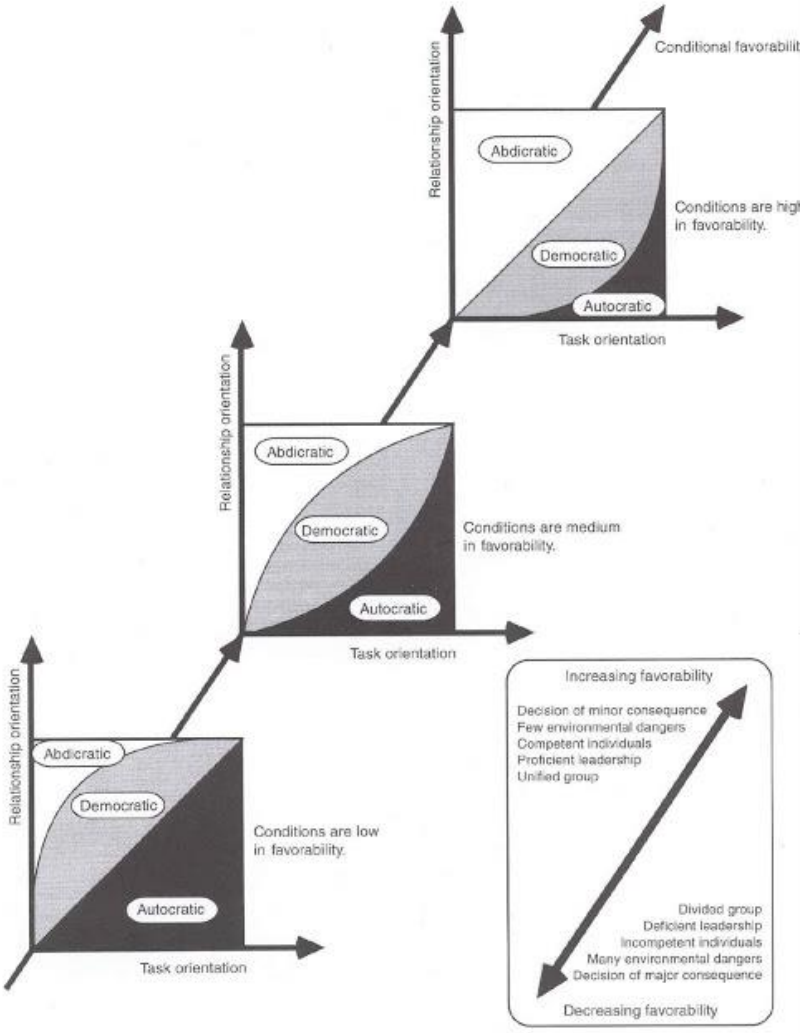
Jordan (1989) argued that leadership theory developed for business and management is inadequate in an outdoor education context and noted that outdoor leadership theory must account for the outdoor setting, perception of risk, and extended leadership periods unique to the profession. From a review of twelve leadership theories organized into four categories, Jordan (1989) proposed the Comprehensive-Interaction-Expectation model (C-I-E). According to Jordan (1989), "Of the four, situational theories appear to suit the needs of outdoor leadership theory more adequately than other theory-bases" (p. 39). Jordan's model attributed leadership to the acceptance and reinforcement

of the leader's roles by the group. Furthermore, Jordan’s model recognized the critical role of the situation.

Priest and Dixon (1991) advanced the Conditional Outdoor Leadership Theory (COLT). The COLT (Figure 1) identified three leadership styles: autocratic, democratic, and abdicratic. Abdicratic leadership involves turning-over all decision-making power to the group (Priest & Gass, 2018). The COLT recognized leadership orientation as task or relationship focused. Finally, the COLT considered conditional favorability. Environmental dangers, individual competence, group unity, leader proficiency, and consequences of the decision are factors in conditional favorability (Priest & Gass, 2018). Priest and Gass (2018) suggested that conditional favorability is the most influential orientation for outdoor leaders. Their model included high, medium, and low favorability. The COLT combines leadership style, leadership orientation, and conditional favorability to form matrices from which a leadership style may be selected.

Figure 1

Conditional Outdoor Leadership Theory



Note. Figure source Priest & Gass, 2005

Brymer and Gray (2006) noted that the most effective blend of Priest and Gass' 12 core competencies is unknown and questioned whether they are a sufficient theory of leadership. They believed that an understanding of transformational leadership must be added for outdoor leaders to be truly effective (Brymer & Gray, 2006). Drawing on work by McKenzie (2000), Brymer and Gray (2006) argued that factors such as facilitating

personal growth among followers demands more than a certain set of competencies. They saw transformational leadership as necessary for outdoor leaders to "consider their own and others values, beliefs, and other personal qualities" (p 14).

Smith and Penny (2010) described three levels of leadership (effective, exemplary, extraordinary) and how they progressively incorporate existing leadership theories. Effective leaders draw upon aspects of situational leadership theory (Hersey et al. 1996), conditional outdoor leadership theory (Priest & Chase, 1989), and transformational leadership theory (Bass, 1999) (Smith & Penny, 2010). Exemplary leadership has been described with components of both transformational and authentic leadership theory (Smith & Penny, 2010). According to Smith and Penny (2010), exemplary leadership is more sophisticated than effective outdoor leadership because it involves the use of the head and the heart.

Finally, Smith and Penny (2010) argued that extraordinary leaders add spiritual leadership to transformational and authentic leadership. They contended that exemplary outdoor leaders use Situational and Conditional Leadership theory as well as draw on concepts related to spiritual leadership. For example, they demonstrated a commitment to experiences as collaborative learning journeys, a sense of calling, and commitment to a greater purpose. Smith and Penny (2010) argued that leadership at this level is a more holistic approach centered on the integration of mind, body, heart, and spirit.

Teaching and Measuring Leadership in Outdoor Education

Chapter one established that leadership development is viewed as a principal outcome of outdoor education programs (Ewert & Overholt, 2010), the importance of demonstrating outcomes of outdoor education programs (Andre et al., 2017), and that

more work needs to be done to establish that outdoor education programs improve leadership skills (Ewert & Overholt, 2010). A review of the existing literature on leadership development in outdoor education follows. The review begins with studies that investigated development of leadership self-efficacy, followed by several studies that explored transformational leadership development. Finally, literature that focuses on leadership more broadly is summarized.

Leadership Self-Efficacy

Self-efficacy is a belief in one's self-capacity for effective performance in a domain-specific endeavor, in this case leadership (Bandura, 1977). Scholars of leadership have identified it as an important construct (Anderson, et al., 2008; Hannah, et al., 2008). According to Propst and Koesler (1998), "Since competency, efficacy, and judgment are considered important prerequisites for leadership (Cain & McAvoy, 1990; Ford & Blanchard, 1985; Petzoldt, 1984), the theory of self-efficacy holds much promise for understanding the outdoor leadership development process" (p. 320). Four studies have examined leadership self-efficacy in programs that use adventure methodology.

Kass and Grandzol (2012) used a quasi-experimental design to examine leadership development of students enrolled in an Organizational Behavior MBA course. The first group participated in a section of the course which included an adventure-based development program referred to as Leadership on the Edge (LOTE). The comparison group completed a section of the course without LOTE. However, it is important to note that this course was taught in an experiential as opposed to conventional classroom format. The same instructor taught both sections of the course. The coverage, sequence, and depth of content was nearly identical for both sections.

Kass and Grandzol (2012) measured generalized self-efficacy, emotional intelligence, and leadership motivation. Leadership motivation was measured with a 14-item scale (Pierce & Newstrom, 2003) designed to measure overall desire and readiness to engage in the leadership behaviors of coaching, building cohesive teams, managing conflict, and influencing. They discovered that both groups showed increases in emotional intelligence, self-efficacy, and leadership motivation over the course of the semester, but the LOTE did not show greater increases than the classroom section, as they had anticipated (Kass & Grandzol, 2012). They explained the results by noting that there was an absence of a true control group. The LOTE course was effective in this circumstance, but the experiential classroom was equally effective from a statistical perspective (Kass & Grandzol, 2012). They concluded by noting that future research on outdoor training initiatives needs to use designs which incorporate a control group so that the added benefits of such approaches can be determined (Kass & Grandzol, 2012).

Two studies (Starbuck & Bell, 2017; Fields, 2010) focused on leaders in outdoor orientation programs. Fields (as cited in Starbuck & Bell, 2017) conducted a sequential explanatory mixed-methods study with 15 outdoor orientation peer-leaders and concluded that students' training and leadership experience increased leadership self-efficacy. Starbuck and Bell (2017) interviewed 36 first-time OOP peer-leaders at four colleges and found that students placed high value on their leadership experiences and gained greater confidence in their leadership ability. They postulated that students had rich experiences due to stage-environment fit:

> If students are seeking leadership opportunities with increased responsibility, but not afforded these opportunities, the mismatch in stage (needing autonomy) and

> environment (restricted from autonomy) could result in frustration and a lack of meeting developmental goals; conversely, if students are seeking leadership opportunities and are afforded responsibilities matching their developmental stage, students would be expected to have successful experiences. (p. 282)

One study (Huey et al., 2014) compared programs using adventure with more conventional leadership training at the United States Naval Academy. Leadership self-efficacy was measured using a 17-item scale, based on a leadership self-efficacy taxonomy (Anderson et al., 2008 as cited in Huey et al., 2014). They found that adventure-based training programs may be more effective than traditional leadership training programs in increasing leadership self-efficacy. The researchers compared outcomes of a group that used adventure-based training methods against a group using a conventional leadership training program at the U.S. Naval Academy. The adventure-based groups participated in an expedition with the National Outdoor Leadership School (NOLS) and a sailing program known as the Off-Shore Sail Training Squadron (OSTS). The conventional training program is known as Plebe Detail. They found that midshipmen at the United States Naval Academy who enrolled in adventure-based training programs experienced greater self-efficacy gains than those participating in a traditional leadership training program. They explained this difference by noting the adventure-based training had greater novelty of the training environment and more instructor-participant feedback.

In summary, these studies suggested that adventure-based training programs increase leadership self-efficacy and in the case of Huey et al. (2014) may be more effective than more conventional training methods. However, belief in one's ability does

not always correlate with actual ability (Huey, 2014). The present study aims to facilitate research beyond development of leadership self-efficacy by creating an instrument to measure actual leadership behaviors.

Transformational Leadership

Brymer et al. (2010) and Hayashi and Ewert (2006) examined transformational leadership within the outdoor leadership context. Both studies compared samples of outdoor leaders with the general population using the Multifactor Leadership Questionnaire (MJQ; Bass & Avolio, 1997; as cited in Hayashi & Ewert, 2006). Hayashi and Ewert (2006) examined responses from 46 (male n = 28, female n = 18) participants. The researchers found that outdoor leaders demonstrated a more transformational style than the general population on two subscales of the MLQ: the way they inspired and motivated students, and the level at which they considered individual student issues. Furthermore, the authors noted that outdoor leaders provided fewer contingent rewards and intervened with corrective behavior only when absolutely necessary. Brymer et al. (2010) used results from 104 completed surveys (male n = 70, female n=34). Brymer et al. (2010) replicated Hayashi and Ewert's (2006) findings that outdoor leaders have higher levels of transformational leadership than the general population and value contingency rewards less.

Both studies were exploratory and descriptive in nature and concluded that future studies should investigate whether outdoor education develops these skills. This conclusion lends further support to the present study's assertion that an authentic assessment of leadership skills is necessary to facilitate studies of leadership development by outdoor programs.

Leadership Generally

Ewert and Overholt (2010) used a modified version of the Empowering Leadership Questionnaire (Arnold et al., 2000) and the leadership section of the Outward-Bound Outcomes Instrument (Frankel & Ewert, 2009) to answer the question of whether outdoor education develops leadership skills. The researcher's sample included college students enrolled in a semester-long outdoor leadership program (male n = 11, female n = 7) or a required class (male n = 36, female n = 47). The participants in the required class served as a comparison group. The study's focal point was a three-week expedition with data collected two days before, three days after, and two weeks after the expedition.

The results indicated a significant increase on both leadership measures after controlling for pretest scores for the entire sample. Increases in the treatment group reported greater gains in leadership development than the control group. Ewert and Overholt (2010) concluded that outdoor leadership training programs can be effective in developing leadership skills. Ewert and Overholt (2010) noted that the study used self-reported leadership skills, not actual leader behaviors, further mentioning that actual leader behaviors may be an important area of focus for future studies.

Bobilya et. al., (2017) used a mixed methods design to investigate outcomes of participating in an North Carolina Outward Bound School (NCOBS) course. The authors used the North Carolina Outward Bound Course Impression Survey (NCOBSCIS) to explore participants' perceptions of the difference of their own leadership development (among other outcomes) prior to and immediately after their participation in an NCOBS course. Bobilya et al's., (2017) sample (male n=172, female n=94) was drawn from participants in an NCOBS open-enrollment wilderness course. They found statistically

significant increases in perceptions of leadership development. These findings were not impacted by gender, but course length was a factor. Bobilya et al., (2017) confirmed Ewert and Overholt's (2010) findings. Like Ewert and Overholt (2010), Bobilya et al's., (2017) work measured self-perception of leadership rather than leader behaviors. The present study aims to address the need for measures of actual leadership behaviors.

Situational Leadership

The present study uses the situational leadership model (SLM) as its focus for several reasons. First, the focus of most outdoor education programs is teaching leadership to young people inexperienced in leadership. Situational leadership is useful as a starting point for discussing leader-subordinate social interaction and the idea of modifying leader behavior to match subordinate attributes (Thompson & Vecchio, 2009). Furthermore, the SLM is easily understood, widely applicable to various leadership situations, and has stood the test of time in leadership training programs (Thompson & Vecchio, 2009). Finally, the SLM is clearly prescriptive in nature and offers guidelines for interpersonal relations (Thompson & Vecchio, 2009). By comparison many other leadership theories are more descriptive. Budding outdoor leaders benefit from the prescriptive nature of situational leadership.

Another important reason for centering this study on situational leadership is that the model has been embraced and adopted in the field of outdoor leadership (Peart, 1991; Phipps & Swiderski, 1990; Ford & Blanchard, 1985). The following section demonstrates the influence of Situational Leadership on outdoor leadership texts, in outdoor leadership training organizations, and in outdoor leadership theory development.

SLM's influence is seen in much of the literature previously reviewed (Jordan, 1989; Priest & Dixon, 1991; Priest, 1986; Smith & Penny, 2010). Furthermore, the SLM is prominently featured in several outdoor leadership textbooks including *Administration and Leadership of Outdoor Pursuits* (Ford & Blanchard., 1985), *The Backcountry Classroom: Lesson Plans for Teaching in the Wilderness* (Drury, et al., 2005), *Teaching Adventure Education Theory* (Stremba & Bisson., 2009), *The Wilderness Educator: The Wilderness Education Association's Curriculum Guide* (Cockrell, 1991) and *Adventure Education* (Miles & Priest, 1990).

Situational leadership is recognized as an important theory used by two prominent outdoor leadership training organizations. The National Outdoor Leadership School (NOLS) uses the 4/7/1 leadership education model (Gookin & Leach, 2009). The model teaches four roles: designated leadership, peer leadership, active followership, and self-leadership. It includes seven leadership skills: vision and action, expedition behavior, communication, competence, tolerance for adversity and uncertainty, judgement and decision-making, and self-awareness. Finally, NOLS teaches that students should develop their one signature style. NOLS teaches leadership as "situationally appropriate action that directs or guides your group to set and achieve goals." Situational leadership is prominently featured as a theory instructors can use to help teach decision-making to students.

The WEA, founded in 1977 was rooted in university outdoor recreation academic programs. It was conceived as a means to train and certify leaders in judgment and decision-making skills to safely lead people on outdoor trips (Berman & Teeters, 2003). At the time of this writing, the WEA is in the process of re-establishing its 6 + 1

curriculum, with an emphasis on situational leadership. Historically the WEA included situational leadership in its curriculum (Phipps, 1991). The WEA emphasized the usefulness of situational leadership when working with groups in expedition settings (Phipps, 1991).

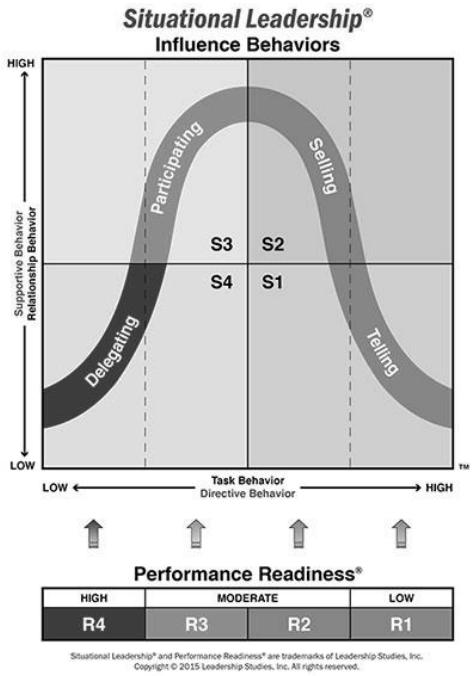
Priest and Dixon's (1991) Conditional Outdoor Leadership Theory (COLT) is prominent in the field of outdoor education. The COLT is highlighted in three commonly used textbooks (Drury et al., 1985; Ford & Blanchard, 1987; Priest & Gass, 2018). The COLT and SLM are similar in that they are both contingency theories. Both rely on leadership orientation (task vs. relationship; Stogdill et al. 1957) and the Tannenbaum-Schmidt (1986) continuum of leader behavior. However, Priest and Dixen, (1991) argue that conditional favorability is more influential in determining leadership style than task vs. relationship. The COLT's individual competence factor is a direct application of performance readiness found in the SLM. The COLT adds environmental dangers, group unity, leader proficiency, and decision consequences as factors. A significant difference between the COLT and SLM is in leadership style. The COLT recognizes three styles (autocratic, democratic, abdicratic), whereas the SLM advances four styles (telling, selling, participating, delegating). This is significant in that some have argued that an abdicratic style is inappropriate in outdoor leadership.

A review of the literature demonstrates the important role situational leadership plays in outdoor education leadership theory. Situational leadership is prominently featured in key outdoor leadership textbooks, integrated into new theories, and utilized by well-respected outdoor leadership training organizations. In the following section methods for teaching and measuring leadership in the outdoor context are explored.

Situational leadership was developed by Paul Hersey and Kenneth Blanchard at the Center for Leadership Studies in the late 1960's. The duo worked on the model together until 1982, at which time Blanchard parted ways with Hersey to work on a modified version of the model called SLII (Hersey et al., 2013). The present study uses SLII (Figure 2) as it is the most current version of the model and the model used in current research (e.g., Thomspon & Glaso, 2018).

Figure 2

Situational Leadership Model



Note. Figure source Hersey et al., 2013

Situational leadership's foundation is the interaction among the amount of guidance and direction (task behavior), the amount of socioemotional support (relationship behavior) a leader provides, and the performance readiness of the followers for a specific activity, task, or job (Hersey et al., 2013). The authors define leadership style as the behavior of the leader as perceived by the followers. Important to the theory

is the identification of task and relationship behaviors (Fiedler, 1967 as cited in Hersey et al., 2013). According to Hersey and Blanchard (2013):

> Task behavior is defined as the extent to which the leader engages in spelling out the duties and responsibilities of an individual or group. These behaviors include telling people what to do, how to do it, when to do it, and who is to do it. Relationship behavior is defined as the extent to which the leader engages in two-way or multi-party communication. The behaviors include listening, facilitating, and explaining the why's of something while offering supportive behaviors to others. (p. 115)

Task behavior is a one-way communication process from the leader to the follower with goal achievement, not feelings of primary concern (Hersey, et al., 2013). Teaching complex outdoor skills such as belaying, navigating during a storm, or responding to an emergency are examples of situations in which a leader is expected to exert high levels of task behavior. By contrast, scenarios characterized by low risk with a competent group suggest greater attention to relationship behavior by the leader. For example, an experienced group planning a route for a day of backcountry travel is low risk. In this example, the leader should engage in relationship-oriented behaviors such as coaching, facilitating, and goal clarification. Using task and relationship dimensions the leader may identify leadership styles.

The Situational Leadership Model includes four basic leadership styles. The appropriateness of the style depends on the situation (Hersey, et al., 2013). Hersey, et al. (2013) define the styles in the following manner: Style 1 (S1) is a leadership style characterized by above-average amounts of task behavior and below-average amounts of

relationship behavior. Style 2 (S2) is a leadership style characterized by above-average amounts of both task and relationship behavior. Style 3 (S3) is a style characterized by above-average amounts of relationship behavior and below-average amounts of task behavior. Style 4 (S4) is a style characterized by below-average amounts of both relationship and task behavior.

The final component of the SLM is the performance readiness of the followers or group. Per Hersey et al. (2013) performance readiness is defined as, "the extent to which a follower demonstrates the ability and willingness to accomplish a specific task" (p. 118). Ability and willingness are the two major components of performance readiness. Ability is the demonstrated knowledge, experience, and skill that an individual or group brings to a task or activity. Hersey et al. (2013) define each as follows: knowledge is demonstrated understanding of a task, skill is demonstrated proficiency in a task, experience is demonstrated ability gained from performing a task.

Individual and group ability is task-specific (Hersey et al., 2013). For example, a group that has demonstrated high ability in backcountry navigation during the backpacking element of a course may possess low ability in the canoeing section of the course. The specific outcome desired must determine judgments of individual or group ability (Hersey et al., 2013).

Hersey et al. (2013) explain willingness as, "the extent to which an individual or group has demonstrated confidence, commitment, and motivation to accomplish a specific task" (p. 118). They are defined as follows: confidence is demonstrated self-assurance in the ability to perform a task, commitment is demonstrated as dedication to perform a task, motivation is demonstrated desire to perform a task.

Hersey et al. (2013) divided performance readiness into four levels. Level 1 (R1). Unable and insecure. The follower is unable and lacks confidence. Or Unable and unwilling. The follower is unable and lacks commitment and motivation. Level 2 (R2). Unable but willing. The follower lacks ability but is motivated or Unable but confident. The follower lacks ability but is confident as long as the leader is there to provide guidance. Level 3 (R3). Able but insecure. The follower can perform the task but is insecure or apprehensive about doing it alone. Or able but unwilling. The follower has the ability to perform the task but is not willing or unmotivated. Level 4 (R4). Able and willing. The follower has the ability to perform and is committed. And able and confident. The follower has the ability to perform and is confident about doing it.

Some of the research on situational leadership has been critical of the model and offered only limited empirical support (e.g., Fernandez & Vecchio, 1997; Thompson and Vecchio, 2009; Vecchio et al., 2006). However, recently Thompson and Glaso (2018) conducted a study with the largest sample (168 supervisors, 830 employees) yet used in an empirical test of situational leadership. They found support for situational leadership principles when leader rating and follower self-rating are congruent. Thompson and Glaso's (2018) results contradicted Fernandez and Vecchio (1997) and Thompson & Vecchio (2009). The authors noted that these studies applied only supervisor ratings of follower's developmental level, whereas they used both leader and follower ratings. Thompson and Glaso (2018) concluded that Situational Leadership principles are more likely to hold when there is congruence between leader and follower ratings of performance readiness.

Thompson and Vecchio (2009) argued that situational leadership should not be completely abandoned by leadership training programs. Thompson and Vecchio (2009) commented, “Clearly, the portion of SLT that has received the greatest support (i.e., low consideration and high structuring for newer recruits) should be taught, and the broader notion of coupling greater autonomy with subordinate experience should be noted as well” (p. 846). These aspects of situational leadership make it well suited to outdoor leadership training due to the fact that most outdoor leadership contexts involve leading groups of novice individuals in an environment characterized by highly technical tasks (e.g., rock-climbing, map reading). In fact, the dynamic of novice followers and discrete, technical tasks found in most outdoor leadership may make diagnosing followers developmental readiness levels more straight-forward than in contexts in which much of the research on situational leadership has been conducted (e.g., business). This point addresses a major criticism of the SLM, that is the ambiguity around the conceptual definition of follower development level (Thompson & Vecchio, 2009). Thus, outdoor leadership may prove to be a useful context to test situational leadership.

Phipps (1986) developed a method to systematically teach outdoor leadership using situational leadership called Experiential Leadership Education (ELE). The ELE combines situational leadership, Jones’ (1973) model of group development, and a group dynamics teaching model (Phipps, 1987). The ELE uses a structured journal technique, the Expedition Leadership Styles Inventory (ELSA; see Grube et al., 2002) and the Group Dynamics Questionnaire (GDQ).

Grube et al. (2002) used a single-subject experimental design to investigate the ELE methodology on an eight-day outdoor leadership course. The ELSA was used as a

pre-test and post-test to compare changes in leadership styles over the course. Grube et al. (2002) used the systematic journal technique developed by Phipps (1986) with the addition of individual student conferencing designed to teach the skill of selecting leadership styles according to the SLM. The researchers reported an increase in leadership effectiveness scores (ELSA) from 75% to 92%. Grube et al. (2002) concluded that the ELE journal technique was effective in teaching the SLM by forcing students to consider the model for each decision and by providing a means for instructors to "get inside their students' heads."

Gabriel (2015) used the ELSA to explore the impact of formal staff training and field leadership experience on the ability of students to discern appropriate leadership for specific outdoor recreation situations. Gabriel (2015) conducted a three-part study with 106 student outdoor leaders from several midwestern United States university outdoor recreation programs. The ELSA was administered prior to formal staff training, at the conclusion of staff training, and after a minimum of seven days of field leadership experience. Gabriel (2015) found no significant differences in ELSA scores from pre-test to formal staff training or field leadership experience. Gabriel (2015) concluded that the studies' failure to consider the non-standardized nature of leadership training and field experiences and low statistical power may have contributed to the results. The preceding section reviewed the relevant leadership literature. In the following section the concept of validity is explored.

Validity

The Standards (2014) state, "validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. Validity is,

therefore, the most fundamental consideration in developing tests and evaluating tests" (p. 11). Despite the importance of validity, the preceding definition does not enjoy universal acceptance by scholars (Bandalos, 2018). However, this definition, along with the unified view of validity will be accepted for this study. The unified view is now widely held, "validity is a unitary concept. It is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed use" (AERA/APA/NCME, 2014, p. 14). The unified view refers to forms of validity evidence, rather than distinct types of validity.

This study proposes to advance scholarship by designing and collecting evidence of validity for a measure of leadership behaviors based on the Situational Leadership Model. Evidence based on content is concerned with the degree to which test content provides an adequate representation of the domain to be measured (Bandalos, 2018). In this case the domain to be measured is leadership as interpreted by the SLM. Generally, it is not feasible to include every item that is part of a construct's domain. This certainly applies to leadership. Thus, the concern in this instance will be the degree to which the rubric items contain a representative sample from the construct of situational leadership. The genesis of evidence based on test content must be a detailed definition of the domain (Bandalos, 2018). The first task, therefore, is to develop clear learning outcomes. It is critical to determine the knowledge, skills, and other attributes to be revealed by the assessment task (Messick, 1995). Matching student learning outcomes or standards to test content is known as alignment.

There are two threats to validity generally that are particularly associated with evidence based on test content: construct underrepresentation and construct-irrelevant

variance (Bandalos, 2018; Messick, 1989). Construct underrepresentation occurs when the instrument is too narrow and does not include important dimensions of the construct (Messick, 1989).

Construct underrepresentation can occur in at least two ways. First, if the rubric samples too narrowly from the construct. Omitting items related to evaluating follower's performance-readiness from the rubric would constitute construct underrepresentation. Moskal and Leydens (2000) suggest that emphasizing items in the rubric unrelated to the construct is a second way in which rubrics may produce construct underrepresentation. For example, if the rubric emphasized spelling, grammar, and sentence structure the student displays in a journal at the expense of leadership knowledge, construct underrepresentation has occurred.

Construct-irrelevant variance occurs when the assessment is influenced by factors not part of the intended construct (Bandalos, 2018). For example, suppose that leadership journal entries are used to score a student's knowledge of leadership. If the journal prompt uses overly complex language, the student's scores may suffer due to their unfamiliarity with the vocabulary, not their understanding of leadership. A discussion of the important elements to be included in a rubric designed to measure situational leadership follows.

Effective situational leaders must possess knowledge of three aspects of the SLM: concepts of supportive and directive behaviors, four basic leadership styles, and performance readiness of followers. Effective situational leaders also demonstrate several abilities. First, they should be able to correctly assess follower's performance readiness. Next, they must be able to utilize each of the four basic leadership styles. Finally,

students have to match the follower's performance readiness with the desired leadership style.

Hersey and Blanchard designed the Leader Effectiveness and Adaptability Description (LEAD) questionnaire for assessing leadership style. LEAD Self was developed to measure leaders' perceptions of their own style, style range, and adaptability. According to Hersey et al. (2013) the LEAD instruments were designed for training purposes, not as research instruments. LEAD has been sharply critiqued for its lack of validity and reliability evidence (Thompson & Glaso, 2018).

Performance readiness has been measured using subjective measures for example, follower self-report or peer/leader assessment. Objective indices like education level and years of experience (Vecchio & Boatwright, 2002) and job level (Fernandez & Vecchio, 1997) have also been used. The manager rating scale and the staff rating scale were both designed to measure job performance readiness (ability) and psychological performance readiness (willingness) on five behavioral dimensions (Hersey et al., 2013). Thompson and Glasso (2018) used a modified ten-item Employee Readiness Scale (Fernandez & Vecchio, 1997).

Phipps' (1996) ELE technique evaluated knowledge of these elements in a structured journal. The journal asks students to select the follower's readiness level, the corresponding style (as suggested by the model), the style actually used, and whether the decision was task, relationship, or both. Phipps evaluated student's ability to match leadership styles and performance readiness with the ELSA. The ELSA uses twelve situations, each with four alternatives representing the four leadership styles found in

situational leadership. The ELSA is intended to function as ‘Johari Window’ in that it calls attention to dominant and supportive styles (Phipps & Phipps, 2003).

Conclusion

The literature reviewed began with classic studies which largely focused on identifying the competencies of outdoor leaders through consensus of outdoor professionals. This research formed an important foundation. It begs the question of how to measure leadership competencies in outdoor leaders? The next generation of research went beyond lists of skill-based competencies to propose new theories incorporating existing leadership theory. A theme of this body of work is the incorporation of contingency theories, especially situational. Next, scholars turned to measure leadership in outdoor programs. Leadership self-efficacy, transformational leadership, and leadership defined broadly measured with self-report instruments are common. A conclusion of this scholarship is that research should continue to investigate the ability of outdoor programs to develop actual leader behaviors (as opposed to self-reported leadership skills). The present study proposes to develop a measure of actual leader behaviors using journaling and a rubric.

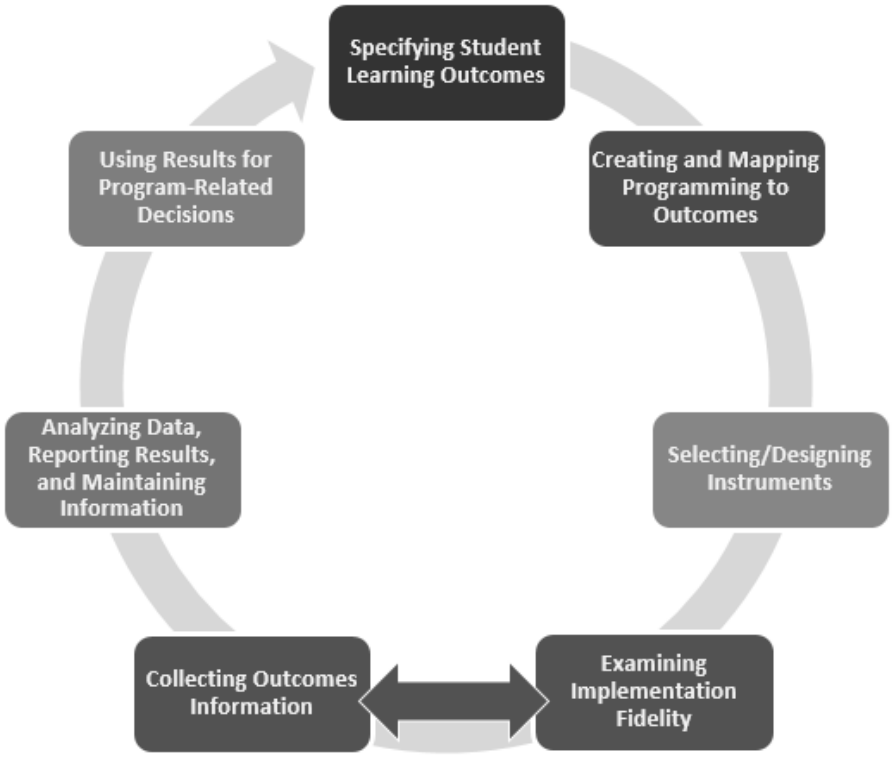
Chapter Three

Methods

The process for this study and the components of the OSLR will be matched with the assessment cycle (see Figure 3): (1) specifying student learning outcomes; (2) creating and mapping programming to outcomes; (3) selecting/designing instruments; (4) examining implementation fidelity and collecting outcomes information; (5) analyzing data, reporting results, and maintaining information; (6) using results for program-related decisions.

Figure 3

The Assessment Cycle



Note. Figure source James Madison University

The learning outcomes were established by a group of three subject matter experts serving on the WEA Standards Working Group. The outcomes were tied to the WEA 6+1 curriculum. For a summary see Table 1.

Table 1

Learning outcomes, journal prompts, and rubric evaluative criteria

Learning Outcome	Journal Prompt	OSLR Rubric
When presented with a problem in the field, identify a leadership style that is appropriate to the group's level of readiness.	Select the **style** corresponding to the group's readiness level according to Situational Leadership. Describe how you implemented the selected style using verb indicators.	Leadership Application
Explain ability as a function of a group's knowledge, skills, and experience. Explain willingness as a function of a group's confidence, commitment, and motivation.	Select the **readiness** level of the group for this decision:	Ability Willingness
When presented with a problem in the field, demonstrate a leadership style that is appropriate to the group's level of readiness.	What style did you use to make this decision? Circle your choice.	Leadership Application

To introduce the proposed methods for this study an example of how a participant will experience the study is presented here. First, a student will participate in an outdoor

leadership training course that includes instruction in Situational Leadership. Second, the student will apply what they have learned by leading a group of peers during a backcountry expedition (e.g., canoeing, backpacking). This takes place through a common practice known as leader of the day. After completing the leader of the day experience the student will record their four most important decisions and reflect on them within the context of Situational Leadership through a structured journal. These journal entries will be scored by trained raters using a rubric designed for this study. Finally, the rubric scores will be analyzed in a g-study. A more detailed explanation of the methods follows.

Participants

Messages were posted to The Association of Outdoor Recreation and Education's (AORE) listserv and the Benefits of Outdoor Education, Research, Studies, and Discussion Resources Facebook page seeking program administrators willing to participate in the study. In addition, the researcher contacted program directors directly with a request to take part in the research. Program directors invited participants to complete the study. The sample (n = 51) consisted of participants from three types of programs: an academic outdoor leadership course situated in a private university in the Western United States (1), campus recreation trip leader programs located within a small private university in the South East (1), and within mid-size state universities in the South East and Western United States (3), finally, an internship program conducted by a non-profit outdoor organization in the North East (1).

The study included 51 (21 male, 19 female, 6 non-binary/third gender, 1 prefer not to say, 4 did not answer) participants in outdoor leadership training programs between

the ages of 18 to 26 (M = 20.60, SD = 1.47). The participants' racial identities were: 41 White; two Black or African American; one Native Hawaiian or Pacific Islander; one multiple races; one some other race, ethnicity, or origin; and five did not answer.

Table 2

Sociodemographic Characteristics of Participants

	n	%
Gender		
Female	19	37
Male	21	41
Nonbinary/third gender	6	12
Prefer not to say	1	2
Did not answer	4	8
Race		
White	41	80
Black or African American	2	4
Native Hawaiian or Pacific Islander	1	2
Multiple Races	1	2
Some other race, ethnicity, or origin	1	2
Did not answer	5	10

Procedures

All program directors were familiar with Situational Leadership. However, they received instruction in Situational Leadership from the researcher to ensure programming

was mapped to the learning outcome (step 2 of the assessment cycle). The instruction included a presentation on situational leadership, and instruction in the use of the structured journal technique. The program directors in turn taught Situational Leadership to the participants using the provided materials.

Course participants were instructed to keep a structured journal (modified from Phipps, 2002, Appendix C) during the duration of the field-based portion of the course. The instructions included: (a) You will complete your journal following your leader of the day experience (b) Record one entry for each of the three most important **leader** decisions of the day, (c) The journal entry should reflect on your own decision-making, not critique someone else's decision-making. Leader decisions were defined as those that "affect someone else or the group. Leader decisions are not personal decisions, like deciding when to put on an extra layer."

Raters

Three professors (2 male, 1 female) of Outdoor Education/Recreation were selected as raters. The raters were selected based on the following criteria: current or former status as a professor of outdoor leadership at an accredited college or university, familiarity with the Situational Leadership Model, and experience training students in outdoor leadership.

Rater Training

The researcher met via Zoom with the raters to conduct training on four occasions. The first meeting included an overview of the project, a review of the Situational Leadership Model, and an explanation of the scoring rubric. After the first meeting, a practice scoring exercise was provided to the raters. The rubric was modified

based on rater feedback following this exercise. Next, following best practice, raters individually scored three example journals designed by the researcher as anchors (see Johnson et al., 2009) representing poor, average, and excellent examples. The raters met to discuss discrepancies in the scores until a consensus was reached on all scores. This discussion served to calibrate raters and ensure each criterion was interpreted in a consistent manner (Johnson et al., 2009).

The raters were provided with a manual for the OSLR which included an overview of SLT and detailed instructions on the use of the OSLR with examples. In addition, raters were provided with the educational materials provided to program directors. Finally, they were given access to the practice scoring exercises completed during the rater training.

Measures

The Outdoor Situational Leadership Rubric (OSLR) was the focus of this study. The rubric was designed to measure students' knowledge of four basic leadership styles put forth in the Situational Leadership Model, and the ability to assess the performance readiness of followers. Students participating in the outdoor leadership training were required to keep a structured journal (modified from Phipps, 2002, Appendix C) throughout the course. A four-element, behaviorally anchored assessment rubric (the OSLR) was used to evaluate the major components of the Situational Leadership Model using students' structured journals (Appendix B). Raters scored each element on a four-point scale (0-no indicators present, 1-indicator undefined, ambiguous, or unexplored, 2-indicator stated, described some aspects lack clarity, 3- indicator explained clearly and

described comprehensively, delivering all relevant information necessary for full understanding) using the structured journal responses.

One of the three legs upon which the SLM is built is performance readiness. This requires judging the group's ability and willingness, which may be done by observing behavior or asking followers (Hersey et al., 2013) and applying the correct indicator or performance readiness based on the group's demonstrated ability and willingness. The SLM defines ability as a group's knowledge, skill, and experience, and willingness as confidence, commitment, and motivation. It is desirable that students consider all six of these indicators of ability and willingness to select the performance readiness of the group.

Participants are judged on their knowledge of this aspect of performance readiness in criteria one (ability) and two (knowledge) of the OSLR (see Figure 4). The structured journal prompt for ability and knowledge (question e) is, "Describe the behaviors you observed (indicators) of ability (knowledge, skill, experience) and willingness (confidence, commitment, motivation) the group demonstrated that led you to choose the readiness level." Raters are asked to determine how well the participant has explained the group's performance readiness on all six criteria. For example (see Figure 4), a participant may write, "On day five the group was taught navigational skills including map reading, they passed our map test with ease. On day 8 they successfully navigated without the help of instructors." This is an example of a full score because the participant has explained clearly and described comprehensively all three aspects of ability (knowledge, skills, experience) in their answer.

Figure 4

Sample journal entry and scoring for performance readiness- ability.

e. Describe the behaviors (indicators) of ability (confidence, commitment, motivation) and willingness (knowledge, skill, experience) the group demonstrated that led you to choose the readiness level.

On day five the group was **taught navigational skills** including map reading, they **passed our map test with ease**. On day 8 they **successfully navigated** without the help of instructor

		0	**1**	**2**	**3**
		no indicators present.	indicator undefined, ambiguous or unexplored	indicator stated, described some aspects lack clarity.	indicator explained clearly and described comprehensively, delivering all relevant information necessary for full understanding.
Ability *See journal part e*	Knowledge				3
	Skill				3
	Experience				3

Participants' scores on the application of performance readiness are derived from criteria three of the rubric (performance readiness application). In criteria three raters compare participants' answers to question d to question e. Question d is, "Select the **readiness** level of the group for this decision." Participants may choose one of the following options: R4 (Able/Willing), R3 (Able/Unwilling), R2 (Unable/Willing), R1 (Unable/Unwilling). Question e is, "Describe the behaviors you observed (indicators) of ability (knowledge, skill, experience) and willingness (confidence, commitment,

motivation) the group demonstrated that led you to choose the readiness level." Here raters are asked to look for how well the participants' description of the observed behaviors (question e) matches the readiness level they selected (question d) according to the theory. For example (see Figure 5), a participant may answer question e as follows, "On day five the group was taught navigational skills including map reading, they passed our map test with ease. On day 8 they successfully navigated without the help of instructors. The group was anxious about solving the problem on their own but wanted to succeed by themselves. Despite some early confusion in the decision-making process, they kept at it." This description indicates an R3 group (able/unwilling or insecure). The rater must simply look at section d to see if the participant has selected the correct readiness level. If the answer to question e lacks sufficient detail to determine a match the rater scores it a zero (unable to assess).

Figure 5

Sample journal entry and scoring for performance readiness application.

d. Select the **readiness** level of the group for this decision:

R4 (Able/Willing and confident**) R3 (Able/Unwilling or insecure)** R2 (Unable/Willing or Confident)
R1 (Unable/Unwilling or insecure)

e. Describe the behaviors (indicators) of ability (confidence, commitment, motivation) and willingness (knowledge, skill, experience) the group demonstrated that led you to choose the readiness level.

On day five the group **was taught navigational skills** including map reading, they passed **our map test with ease**. On day 8 they **successfully navigated** without the help of instructors. The group was **anxious about solving** the problem on their own but wanted to succeed by themselves. Despite some early confusion in the decision-making process, they kept at it.

Performance Readiness Application	**0** mismatch or unable to assess.	**3** match	
See journal parts d and e		3	

After determining the performance readiness of the group, leaders must apply the appropriate leadership style. Application of Situational Leadership is assessed in criteria four of the OSLR (Leadership Application). Raters start by comparing question b, "What style did you use to make this decision? With the following options: S1 (Telling), S2 (Selling), S3 (Participating), S4 (Delegating) to question e (see above) to assess if the participant has matched leadership style with the group's performance readiness. This is fundamental to the theory, failure to match style and readiness earns a score of zero. For example (see Figure 6), if a group's readiness level is R4 (able/willing and confident) then the leader should choose S4 (delegating).

Figure 6

Sample journal entry illustrating match of leadership style with performance readiness.

b. What style did you use to make this decision? Circle ONE:

S1 (Telling) S2 (Selling) S3 (Participating) **S4 (Delegating)**

d. Select the **readiness** level of the group for this decision. Circle ONE:

R4 (Able/Willing and confident) R3 (Able/Unwilling or insecure) R2 (Unable/Willing or Confident) R1 (Unable/Unwilling or insecure)

If the participant has correctly matched style and leadership the rater can move to question c, "Describe how you implemented the selected style using verb indicators." At this point the rater is judging how well the participant had described the leadership style (see Figure 7). A participant scores a three when they state the style clearly and describe it comprehensively including verbs consistent with the leadership style. These verbs are derived from a list created by the theory's authors and are included in the training materials provided to participants. For example, verbs associated with one end of the leadership spectrum, the telling style, include directing, structuring, and informing. By contrast, verbs describing the other end of the spectrum, delegating, are observing, monitoring, and entrusting.

Figure 7

Sample journal entry and scoring for leadership style application.

b. What style did you use to make this decision? Circle ONE.

S1 (Telling) S2 (Selling) **S3 (Participating)** S4 (Delegating)

c. Describe how you implemented the selected style using verb indicators. We gathered the group and explained that we had **confidence in them** and wanted them to use the skills they had learned over the previous 9 days to consult the map and make the decision. We **shared our ideas and listened** as they discussed what to do, **asked clarifying questions**, and **answered all their questions**. We encouraged them by reminding them of their navigational success on day 8.

		Selects the correct style for the readiness level (compare part b & d) and...		
Leadership Application *See journal part c*	**0** Does not apply the appropriate leadership style for the performance readiness of the group. Style is not correct for readiness level (compare part b & d)	**1** the description of the style is missing, vague, or inconsistent with the leadership style	**2** the description of the style is described, some aspects lack clarity.	**3** provides an explanation of the style stated clearly and described comprehensively, delivering all relevant information necessary for full understanding using verbs (eg. coaching, directing, supporting) which are consistent with the style according to the SLM

Statistical Analysis Outdoor Situational Leadership Rubric

Generalizability theory (G-theory) provides a framework for questioning the generalizability of performance assessments (Linn et al., 1991). According to Bandalos (2018) generalizability theory may be used to gather information on the relative magnitude of different sources of error of a test and produce dependability coefficients. G-theory exposes limits that should be placed on interpretation of test scores (Bandalos, 2018). The following section explains G-theory, describes the sources of error in the proposed study, and explains how a dependability coefficient will be used to address the proposed research question.

In G-theory, sources of errors are known as facets, and their levels as conditions. In this study, raters are an example of a facet. The conditions over which we can generalize is known as the Universe of Admissible Observations (UAO) and includes all the levels of facets of interest (Bandalos, 2018). For example, in this study any outdoor recreation faculty member could possibly serve as a qualified rater. Thus, the UAO is scores from all outdoor recreation faculty.

Facets can be fixed or random. Facets are random when the conditions used in the study are assumed to be interchangeable with any others in the UAO (Bandalos, 2018). In contrast, fixed facets represent all those that the researcher wishes to generalize to or all of the possible conditions in the UAO (Bandalos, 2018). This study assumes that outdoor recreation faculty are “interchangeable” and therefore are treated as random facets. The OSLR was constructed to assess the specific elements of Situational Leadership (assessing performance readiness and application of appropriate leadership style). However, there are multiple ways in which these specific elements could be measured. Thus, in this study, the four elements of the OSLR were considered random.

Generalizability theory will be used to estimate the amount of variance for all possible sources of variance in the Universe of Admissible Observations (UAO). The object of measurement is the facet which the researcher wishes to make inferences about, in this case students (σ_s^2), participating in the outdoor leadership courses. Variance in the facets (e.g., raters, items, journal entries) indicates on average the amount of inconsistency across conditions of the facet. Raters (outdoor education faculty) (σ_r^2), items (σ_i^2), and journal entries (σ_j^2) will be random facets. The rater facet indicates the degree of rater severity (leniency or severity) whereas, the item facet indicates differences in item difficulty, and the journal facet indicates systematic differences in journal entries. Crossed facets are those in which conditions of one facet are observed with all conditions of the other (Bandalos, 2018). Each rater will rate all the student's journals on all the tasks.

The purpose of this study was to explore validity evidence based on the internal structure of the rubric. Scores from the rubric will be used in a G-study to determine the amount of error present in each source. For example, the study will yield information on the amount of variance from various sources including the students, raters, and tasks. Variance due to students' scores represent true score variance, where the true score is defined as the average score if a participant repeated the journal writing exercise hundreds of times (Bandalos, 2018). A large proportion of the total variance subsumed by the object of measurement (students) is desirable because it reflects real differences (Burns, 1998). In contrast, variation associated with the facets (tasks, raters), interactions, and residual is undesirable because they represent error. This kind of information can be used to make meaning from test scores and is relevant to test validity (Bandalos, 2018).

The sources of variance from the G-study will be used to address hypothesis two. G-theory is similar to ANOVA, however G-theory's purpose is estimating the amount of variance due to different sources of error, rather than testing the statistical significance of effects (Bandalos, 2018).

Fundamental to G theory is the concept of dependability. Dependability is, "the extent to which the generalization one makes about a given candidate's universe score based on an observed test score is accurate" (Sawaki, 2012, p. 2). The primary research question for this study is, can outdoor education faculty members dependably rate situational leadership using a newly designed rubric? To answer this question measures of the dependability of the raters scores will be used. These measures are known as the index of dependability or phi coefficient and the generalizability coefficient (G-coefficient). Both measures of dependability will be used because scores could be used for both absolute and relative decisions. Absolute decisions are made in comparison to a standard (criterion referencing). Relative decisions are used when comparing individual scores to a group (norm referencing).

The phi coefficient is appropriate when the interest is absolute, whereas the G coefficient is desirable for relative decisions (Bandalos, 2018). A primary interest of the OSLR is how students' scores compare to the rubric's behavioral anchors (i.e., absolute dependability). However, it may be desirable to compare OSLR scores across two groups that experienced different leadership training methods (i.e., relative dependability). Thus, both measures of dependability will be reported.

The measures of dependability produce scores from 0 – 1, with scores closer to 1 indicating greater dependability. The phi-coefficient considers more sources of error and

is generally lower than G coefficients (Bandalos, 2018). Guidelines for acceptable values for measures of dependability have not been established in the literature. However, because the G-coefficient is conceptually similar to that of reliability in Classical Test Theory the guidelines for coefficient alpha may be used as a frame of reference (C. DeMars, personal communication, February 7, 2022). Thus, .80 will be used as a cutoff for this study.

Chapter 4

Results

A G-study is one type of study included in G-theory, the purpose of which is to gather as much information as possible about the size of the sources of error of interest (Bandalos, 2048). Equation (1) shows the decomposition of the observed score variance for the journal within student-by rater-by item design *(j:s) x r x i* into that due to students (σ_s^2), raters (σ_r^2), and items (σ_i^2). The journal entries are nested in students, that is the entries are associated with, yet unique to each student. Nested facets variance components are combined because they are confounded. Additionally, the equation shows the following interactions: rater-by-student (σ_{rs}^2), rater-by-item (σ_{ri}^2), student-by-item (σ_{si}^2), journal within students $\left(\sigma_{j:i}^2\right)$, and journal-by-item ($\sigma_{ji,}^2$). Finally, the equation displays the three-way interactions and error term.

$$\sigma^2(\delta) = \sigma_s^2 + \sigma_r^2 + \sigma_i^2 + \sigma_{rs}^2 + \sigma_{ri}^2 + \sigma_{si}^2 + \sigma_{sj,j}^2 + \sigma_{rsj,rj}^2 + \sigma_{rsi}^2 + \sigma_{sji,ji}^2 + \sigma_e^2 \quad (1)$$

A G-study of the faculty ratings produced the variance components in Table 3. The object of measurement was students. The student variance component (σ_S^2) in the G-study was .03, indicating that approximately 2% of the total variance in the ratings can be attributed to differences between students. Therefore, hypothesis 2: The largest source of the variance will be subsumed by students σ_S^2 was not supported.

The three facets included in this study were the three raters (σ_r^2), three journal entries (σ_j^2), and eight items (σ_i^2) evaluated by the rubric. The rater facet, ($\sigma_r^2 = .07$) which is the amount of variability due to differences in rater stringency explained 6% of the total variance. The item facet ($\sigma_i^2 = .03$) is the amount of variability due to

systematic differences among items. In other words, are some items harder than others? The item facet explained 2% of the total variance.

The two-way interactions explained 22% of the variance. They are explained here in order of the amount of variance each contributed. The first, journal within students $\left(\sigma^2_{j:i} = .11\right)$ interaction amounted to 9% of the variance, which represents systematic differences in the three journal entries within students. This demonstrates how students vary in which entries they find more difficult. Followed by rater-by-item ($\sigma^2_{ri} = .10$) an indicator of systematic differences in rater stringency by item which explained 8% of the variance. Last was student-by-item (σ^2_{si} = .06) explaining 5% of the variance. This shows how different students do better on different items. The remaining two-way interactions did not account for any variation. The three-way interactions are not useful to interpret and typically thought of as error.

Table 3

Estimated Variance Components for a Three-Facet, Crossed (j:s) x r x i Design.

Component	Estimated variance component	Percentage of total variance
Student	.03	2%
Raters	.08	6%
Item	.03	2%
Rater*Student	.00	0%
Rater*Item	.10	8%
Student*Journal, Journal	.11	9%
Student*Item	.06	5%
Rater*Student*Journal, Rater*Journal	.09	7%
Rater*Student*Item	.00	0%
Student*Journal*Item, Journal*Item	.17	14%
Error	.57	46%
Total	1.23	100%

The measures of dependability range from zero – one, with indices closer to one indicating greater dependability. The concept of dependability replaces that of reliability used in classical test theory (Bandalos, 2018). The measures of dependability were ($G = .30, \Phi = .22$). The faculty raters produced ratings well below the .80 benchmark established for adequate dependability. Hypothesis 1: The G and phi-coefficients for the

scores will be adequately high (> .80) was not supported. I explain each coefficient in more detail next.

The G coefficient is desirable for relative decisions. As seen in (2) the only error considered is that which alters students' relative standing within the group. Thus, the main effects of raters and items are not included in the error term because they impact all scores in the same way. Relative decisions are used when comparing individual scores to a group (norm referencing). For example, how individual students compared to the rest of the group in this sample. The G-coefficient shows how consistent students were rank ordered by raters and items. The low G-coefficient ($G = .30$) indicates that raters did not rank order students consistently.

$$G = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{Rel}^2} \qquad (2)$$

$$\sigma_{rel}^2 = \frac{\sigma_{rs}^2}{n_r} + \frac{\sigma_{ri}^2}{n_{ri}} + \frac{\sigma_{sj,j}^2}{n_j} + \frac{\sigma_{si}^2}{n_i} + \frac{\sigma_{ji}^2}{n_{ji}} + \frac{\sigma_{rsj(rj),rsi,sji(ji),e}^2}{n_r n_j n_i} \quad (3)$$

$$\sigma_{rel}^2 = \frac{0}{3} + \frac{.10}{3(8)} + \frac{.11}{3} + \frac{.06}{8} + \frac{.17}{3(8)} + \frac{.83}{3(3)(8)} = .00 + .00 + .04 + .01 + .01 + .01 = .07 \quad (4)$$

$$G = \frac{.03}{.03 + .11} = \frac{.03}{.11} = .31 \quad (5)$$

Absolute decisions are made in comparison to a standard (criterion referencing). The phi coefficient is appropriate when the interest is absolute. A primary interest of the OSLR is how consistent rank ordering would be if different raters were substituted (i.e., absolute dependability). For example, if an organization chose to use a cut score on the OSLR as part of a certification course, would rank-ordering of students be consistent across different raters and different items? The low phi coefficient ($\Phi = .22$) does not

suggest that rank-ordering of students would remain consistent if multiple different raters used the OSLR to assess students.

$$\Phi = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{ABS}^2} \qquad (6)$$

$$\sigma_{abs}^2 = \frac{\sigma_r^2}{n_r} + \frac{\sigma_i^2}{n_i} + \frac{\sigma_{rs}^2}{n_r} + \frac{\sigma_{ri}^2}{n_{ri}} + \frac{\sigma_{sj,j}^2}{n_i} + \frac{\sigma_{si}^2}{n_{ji}} + \frac{\sigma_{ji}^2}{n_j} + \frac{\sigma_{rsj(rj),rsi,sji(ji),e}^2}{n_r n_j n_i} \quad (7)$$

$$\sigma_{abs}^2 = \frac{.08}{3} + \frac{.03}{8} + \frac{0}{3} + \frac{.10}{3(8)} + \frac{.11}{3} + \frac{.06}{8} + \frac{0}{3(8)} + \frac{.83}{3(3)(8)} = .03 + .00 + .00 + .04 + .01$$

$$+ .01 + .01 + .01 = .10 \qquad (8)$$

$$\Phi = \frac{.03}{.03 + .10} = .22 \quad (9)$$

As 2 and 6 demonstrate the coefficients are ratios of true score variance (σ_S^2) to the expected observed score variance (universe plus error variance; Bandalos, 2018). Thus, the low universe score variance resulted in small coefficients of dependability.

Chapter 5

Discussion

In this section I will discuss the results of this study. First, I will interpret the results of the G-study by variance component. Second, I will offer plausible explanations for the results of the study. Third, I will elaborate on the limitations of the study. Fourth, I will make recommendations for how this study may inform future research. Fifth, I will discuss implications of the findings for the field of leadership.

Interpretation of Results

The strength of a G-study is that it divides variance into parts, allowing researchers to examine multiple sources of error. This study included students as the object of measurement (σ_s^2) and three facets. Recall that facets are sources of error. The facets included the three raters (σ_r^2), three journal entries (σ_j^2), and eight items (σ_i^2) evaluated by the rubric. Next, I interpret each of the facets and their associated interactions and offer recommendations based on these interpretations.

Recall that journal entries were confounded by students and the variance components are combined. The confound comes from the fact that the journal entries are associated with each student and journal entries are different for each student. There was a journal within students $\left(\sigma_{sj,j}^2 = .11\right)$ effect, demonstrating how students varied in which entries they found more difficult within students encompassing 9% of the variance. All three journal entries were identical. Students were asked to describe a situation in which they had to make a decision, then explain how they implemented the principles of Situational Leadership in their decision-making. Because each of the three entries

required the same knowledge students should not have varied from one to the next on which entry they found more difficult. Why would there be differences in the scores from one entry to the next? Perhaps students were able to identify one meaningful decision from their leadership experience and adequately describe how they applied Situational Leadership to that decision; however subsequent decisions were not well defined or explained. Or some decisions were easier to apply situational leadership to? Or did journaling motivation wane from the first to the second or third entry?

Why would there be differences between students based on which of the three journal entries they found more difficult? One explanation is that some students were able to effectively describe one or two decisions while others were not able to describe any effectively, resulting in the interaction observed here. Based on these observations I conclude that students need more instruction and practice in the use of the journal.

The item facet (σ_i^2) represented 2% of the overall variance, demonstrating that students found some items harder than other items. If all of the items equally represent Situational Leadership than the items should not contribute any variance. The student-by-item (σ_{si}^2 = .06) interaction explained 5% of the variance. This shows how different students do better on different items. This is similar to the previous interaction; students should not differ on which item they find easy or difficult. The results do not warrant replacing any items. However, several improvements could be made which may lower variance due to items. First, the data of this study should be examined to determine if patterns of item difficulty can be uncovered. Future, studies should place emphasis on instruction to better prepare students to respond to these items.

The rater-by-item ($\sigma_{ri}^2 = .10$) interaction is an indicator of systematic differences in rater stringency by item which explained 8% of the variance. For example, one rater may have been more stringent on item two which assess the students' ability to describe the groups indicators of willingness than other raters. More attention to rater training is needed to minimize the differentiation of scores from one rater to another by item.

The rater facet ($\sigma_i^2 = .07$) comprised 6% of the total variance. Each item is scored on a zero-three scale, except for item seven which is dichotomously scored zero or three. On average rater one awarded .41 points per item, rater two 1.14, and rater three .71. Missing data contributed to a slight underestimate of rater one's mean score, because missing entries were calculated as zero. Rater two scored more leniently. The interaction of rater-by-student (σ_{rs}^2) was zero. While overall rater scores varied the rank-order of students did not vary by rater. This is important because it suggests the rubric functioned as intended in that raters consistently rank-ordered students.

It seems reasonable to conclude that the rater training, scoring manual, and scoring procedures were adequate, but some improvements may decrease the contribution to the overall variance by raters. For example, in this study the researcher created the journal entries used in the rater training and calibration exercises. Future studies may benefit by using actual student journal entries in the rater training and increasing the number of practice journals.

Hypothesis two of this study was that the object of measurement, students, would represent the largest source of variance in the G-study. The student variance component σ_s^2 in the G-study was .03, indicating that approximately 2% of the total variance in the

ratings are attributed to differences in scores between students (true score). Therefore, hypothesis two of this study was not supported. This may be the most significant factor in this study and is explored more fully in the following section.

There are four plausible explanations for the small student variance component: the raters were overly stringent in their ratings, students were not able to convey their thoughts effectively in writing, students did not invest much effort in the task, or the task was simply too difficult. Each of these is examined in turn.

Student's journaling ability was almost certainly a source of construct irrelevant variance in this study. Three conditions are important for effective reflective journaling: perceived trustworthiness of the journal reader, clarity of the expectation, and quantity and quality of the feedback (Kerka, 1996 as cited in Hubbs & Brand, 2005). Incorporating these conditions into future studies may improve the quality of the journal entries. For example, students in this study did not receive feedback on their journal entries. Allowing for practice entries and quality feedback prior to the final journal entries is likely to improve the end product. This could be achieved by providing excellent and poor sample journal entries coupled with instruction explaining what makes for an excellent journal response.

Most students in this study were participants in co-curricular trip leader training programs. It is possible that they lacked motivation for this task. The present study did not include any indicator of motivation. Future studies should incorporate a measure of motivation to determine if this is factor. Additionally, journals should be used in an academic course, attaching a grade to the journal entries. Increasing students' motivation to put effort into all three journal entries may improve scores.

It may be that the task was too difficult. Students had limited instruction in Situational Leadership and practice with the journal prior to their practical leadership experience. The researcher provided a power-point presentation to each of the research sites and review of the theory with the individual responsible for delivering the training. However, the length or quality of the training was not assessed. It is possible that the training was not adequate for students to grasp the theory. The overall scores support this assertion. The total possible score on the OSLR was 72 points. The scores averaged across the three raters ranged from 4.33 to 32.33 (M = 18.10, SD = 7.53). Further evidence of lack of true score variance comes from item seven of the scoring rubric. Item seven evaluates students' ability to match leadership style (S) with the performance readiness (R) of the group. One needs only a superficial understanding of Situational Leadership to know that S and R must be matched. Yet, students consistently struggled with this task. This suggests that there was very limited true score variance among students in this sample.

Probing the possibility that the task was too difficult would be incomplete without turning a lens on the theory itself. Shortcomings in the theory may explain the low scores obtained by students. For example, determining performance readiness may be far more difficult than the theory suggests. This criticism is supported in the literature, according to Thompson and Glaso (2014), "while leader styles seem fairly well understood, establishing follower need for a specific style of leadership has proven to be more difficult" (p. 528). Possibly compounding this problem is the fact that the theory has been applied to groups, rather than individuals in the Outdoor Education context. Hersey and Blanchard (2013) suggest that leaders may have to behave differently individually with

members of their group from the way they behave with the group as a whole. Navigating this complexity may be too much to ask of budding leaders.

The preceding discussion of the variance components of this study assessed the extent to which the scores were impacted by different sources of measurement error. For example, the scores were unaffected by the interaction of rater-by-journal. On the other hand, raters and the raters-by-item interactions contributed error. This information informs the meaning that can be made by the OSLR test scores and therefore is relevant to validity (Bandalos, 2018). The evidence provided by this study does not support using the OSLR in high-stakes decisions like hiring, promotion, or certification. In contrast, the coefficients of dependability are indicators of reliability and are addressed next.

The research question for this study asked if faculty members could dependably rate Situational Leadership using a newly designed rubric. Results demonstrated that three faculty members could not rate Situational Leadership dependably across eight elements ($G = .30, \Phi = .22$). The faculty raters produced ratings well below the .80 benchmark established for adequate dependability. As previously explained it may be that a lack of true score variance contributed to the low G and Phi coefficients.

Limitations

Like any study this research was subject to limitations. Limits related to methodology, data, and sample are addressed. There were several limitations of the methodology with regard to quantity and quality of the instruction provided for students. First, was the small amount of time allowed for instructing the students in Situational Leadership. Most research sites were able to dedicate limited time to Situational Leadership training as part of a comprehensive training occurring over a period of less

than one week. Furthermore, students were not provided with an opportunity to practice with the structured journal prior to its use in the field. Another challenge is that the researcher was not able to assess the quality of the training delivered to the students. Finally, no assessment of the prior knowledge of those delivering the training took place. Program managers were recruited based on prior experience with Situational Leadership and the researcher reviewed the educational materials with each manager. However, it is possible that lack of knowledge of the construct by the trainers limited the quality of the training provided to students.

A limitation of the data is that one rater provided missing scores for seven total points of data. This is likely due to a confusing method for recording the scores but could have been indicative of a greater misunderstanding of how to apply rubric scoring. Thus, the mean score for this rater is slightly underestimated. Maximum likelihood estimation was utilized in the G-study to address the missing data.

Finally, the scope of the study imposed some limitations. The sample was composed of college students. Thus, it may not be appropriate to use the instrument outside of that population. Lastly, a limitation is that this study addressed only one type of validity evidence, internal structure. Further studies will be warranted to gather other types of validity evidence.

Implications

For Programs. The literature review documented the rapid growth of outdoor leadership training programs situated in non-profits (NOLS, OB, WEA), campus recreation, outdoor orientation programs, and academics. Each of these sectors is challenged to demonstrate how they are meeting their mandate to develop leaders.

However, our ability to measure the leadership outcomes of these programs has not kept pace with their growth.

Campus outdoor recreation programs must communicate the value of their programs to school administration, political decision makers, the media, and their constituents (Andre et al., 2017). Nearly three-quarters of outdoor orientation programs identified enhancing student leadership skills as a goal of their program (Galloway, 2000). Outdoor leadership academic programs are expected to meet employers' expectations for developing leadership skills in college graduates (Seaman et al., 2017). Yet, the criteria upon which judgements are made about leaders lack clarity (Smith & Penny, 2010) and few evaluation instruments exist to evaluate outdoor leader effectiveness (Phipps et al., 2005). This study was an important step in creating an authentic assessment of leadership and a tool to meet the demand for accountability in these programs. At the same time, the results of this research serve as an important cautionary tale.

The findings of this study demonstrate that instruments designed to measure leadership may not be functioning as intended. Organizations like the National Outdoor Leadership School and Outward Bound exist to train outdoor leaders; they should exercise discretion when using untested instruments to evaluate and provide feedback to students on their leadership behaviors. Additionally, decisions about hiring and promotion of staff within these organizations should not be left to instruments that lack rigor. This lesson applies to the world of business as well, where over half of organizations use in-house designed leadership measures (Crawford & Kelder, 2017). Academic programs should dedicate resources to developing leadership assessments with

evidence of validity and reliability so that they can be confident in their evaluations of students' leadership abilities.

Even more caution is warranted for the WEA which aspires to certify outdoor leaders. The use of instruments without validity evidence is at best unwise and at worst negligent. In fact, the Standards for Educational and Psychological Testing mandate evidence of reliability and validity for instruments used for credentialing (AERA/APA/NCME, 2014). The findings of this study should serve as a wake-up call to organizations using un-tested leadership measures.

Situational Leadership has been embraced and adopted in the field of outdoor leadership (Peart, 1991; Phipps & Swiderski, 1990; Ford & Blanchard, 1987), yet lacks a psychometrically strong leadership evaluation tool. The only existing tool, LEAD, has been sharply critiqued for its lack of validity and reliability evidence (Thompson & Glaso, 2018). If outdoor education leadership theory continues to lean heavily on Situational Leadership, then a strong instrument is needed to measure it. For example, NOLS teaches leadership as "situationally appropriate action that directs or guides your group to set and achieve goals." At the present time NOLS does not have a strong tool to measure a student's ability to employ "situationally appropriate action" in an authentic leadership assessment task. Phipps (1986) developed Experiential Leadership Education (ELE) method to teach Situational Leadership. Phipps evaluated student's ability to match leadership styles and performance readiness with the Expedition Leader Style Analysis (ELSA). Phipps' contributions to teaching methodology for Situational Leadership are noteworthy. However, the ELSA is intended to call attention to dominant and supportive styles (Phipps, 2003) not measure Situational Leadership. In fact, Phipps'

(1992) recommend the ELSA be used “mainly as a training instrument to illustrate dominant and alternate styles and effectiveness” (p. 23). The OSLR has the potential to fill an important gap in the ability to measure Situational Leadership.

For Research. I offer recommendations in two areas. First, how the findings of this study can inform future psychometric studies. Second, I elaborate on some questions suggested by this research about Situational Theory more broadly.

Future measurement research should build on the findings from this study to increase the likelihood of finding true score variance. One way this might occur is by increasing the duration and quality of the instruction given to students. The methodology can be improved by adding extrinsic motivation through grades, allocating more time for instruction, and creating an assessment of knowledge gained from the instruction. Using the Expedition Leadership Styles Inventory (ELSA; Phipps & Phipps, 1987) with students is recommended for introducing students to the theory, potentially increasing their knowledge. Additionally, providing all students in the sample with several opportunities to practice with and get formative feedback on the structured journal may improve true score variance. For example, providing students with a case-study and asking them to use the structured journal to respond could aid in proficiency using the journal. Replicating this study utilizing a standard curriculum and assessment instruments in an academic outdoor leadership course is well-suited to this purpose.

This study’s findings suggest that further scrutiny of the conceptual basis of Situational Leadership may be warranted. For example, considerable ambiguity exists around the conceptual definition of follower development level (Thompson & Vecchio, 2009). This is especially true when applied to groups. Situational Leadership theory

makes no distinction between individuals or groups as units of evaluation (Johansen, 1990). I have posited that the dynamic of novice followers and discrete, technical tasks found in most outdoor leadership situations may make diagnosing followers' developmental readiness levels more straight-forward than in contexts in which much of the research on Situational Leadership has been conducted (e.g., business). Research that evaluates this proposition is needed.

It is possible that outdoor education has been overenthusiastic in its embrace of a theory that has been criticized by several researchers (e.g., Fernandez & Vecchio, 1997; Thompson and Vecchio, 2009; Vecchio et al., 2006) for lacking empirical support. Perhaps, before more work is done on measuring the theory, we should take a step back and investigate the validity of the theory in an outdoor education context.

Several important questions need to be investigated for scholars to fully endorse Situational Leadership for the field of outdoor education. First, can leaders reliably and consistently judge followers' readiness level? Scholars are encouraged to consider Thompson and Glaso's (2018) findings that Situational Leadership principles are more likely to hold when there is congruence between leader and follower ratings of performance readiness in this line of inquiry.

Next, the conceptual muddiness in the distinction between individuals or groups as units of evaluation needs to be clarified. Crawford and Kelder (2018) remind us that when designing leadership measures, we must consider the boundary/scope conditions of the construct: answering the question where does this theory apply, or not apply? An answer to this question must be addressed before we can move forward.

Most importantly, as scholars of outdoor education we should design experiments to determine if the fundamental tenets of the theory hold. For example, are better outcomes achieved when leaders match their style to the groups readiness level as the theory suggests? The broader leadership literature has done this (see: Thompson & Vecchio 2009; Thomspon & Glaso, 2018) it is time for a replication in the outdoor education context. Only then can we decide if we should jump "all in" with Situational Leadership theory.

For Outdoor Educators. Despite its limitations and measurement challenges Situational Leadership in part, if not in whole has utility for practitioners. Even the theory's critics have suggested that aspects of the theory such as low consideration, high structuring for novice followers, and increasing autonomy with follower experience have value (Thompson & Vecchio (2009). For example, emphasizing to leaders the importance of providing ample guidance to a group of beginner rock climbers to increase the odds of a successful outing. Until the previously identified questions are answered it may be wise for practitioners to use the theory as a starting point for discussions about decision-making, leadership behavior, and followers' expectations rather than a prescriptive model (Johansen, 1990). Even if Situational Leadership does not hold together as a conceptually whole theory, various pieces of it will be useful for practitioners.

Conclusion

Leadership assessment has challenged contemporary scholars (Crawford & Kelder, 2019); this study reflects these challenges. Scholars are confronted with a universally recognized yet complex construct without a widely accepted definition (Day & Antonakis, 2012) and dizzying array of theories representing a diversity of theoretical

perspectives (Lord & Dinh, 2012). All survey measures of leadership have inherent limitations. Therefore, we need to begin to expand our repertoire of tools to examine leadership, which could include observations, interviews, content coding of materials, etc. (Antonakis, et al., 2003). Despite the limitations of this study, it contributes to the nascent discussion of authentic assessment of leadership in an outdoor education context. The study has provided a blueprint for how journaling and rubrics may be used in an authentic assessment to measure leadership. One day it may contribute to the development of a psychometrically strong instrument that meets the accountability demands of a growing field.

Appendices

Appendix A

Summary of classic outdoor recreation leadership research

Author	Population	Sample N	Sample Demographics Gender	Sample Demographics Race	Methodology	Findings
Buell (1981)	Outdoor adventure professionals in the US and Canada	120	80 male, 40 female 37.5% from the New England Region	Not reported	Survey	Design/use of first aid kit Knowledge of group safety Possess physical fitness Limit activities to capabilities Anticipate problems Provide standard of care Apply physical/emotional care Develop safety procedures
Green (1981)	Outdoor leaders in the Pacific Northwest	61	50 male, 11 female	Not reported	Delphi consensus	Risk management plans Small group dynamics Liability considerations

						Outdoor leadership methods Judgement Minimum impact practices Decision-making Assessment of group capabilities Assessment of individual capabilities Outdoor leadership objectives
Swiderski (1981)	Outdoor leaders in the western United States	148	121 male, 24 female, 3 no response	Not reported	Survey	Exercise good judgement Handle safety problems Prepare for accidents Prevent illness/injury Teach environmental injuries Follow a wilderness ethic Model positive attitudes Demonstrate minimum impact Recognize own limitations Recognize problem indicators
Priest (1984)	Practitioners at the AEE conference.	189	94 male, 95 female	Not reported	Survey	Ability to anticipate accidents

						Wilderness first aid skills Awareness of group dynamics Ability to clearly identify problems Ability to evaluate natural hazards Ability to foster teamwork Ability to provide personal growth Proficiency in land-based activities Proficiency in water-based activities Ability to prepare accident responses
Railoa (1986)	Expert panel	5	4 male, 1 female	Not Reported	Survey, pilot test	Leadership style Judgement (objective/subjective) Trip planning/organization Environmental issues Risk management Instructional principles Navigation Group dynamics Nutrition Field experience
	Students	7				

Priest (1986)	Experts from five nations	169	156 male, 13 female	Not reported	Survey	Motivational philosophy/interest Physical fitness Healthy self-concept and ego Awareness and empathy Personable traits and behavior Flexible leadership style Technical activity skills Safety skills Organizational skills Environmental skills Instructional skills Group-management skills Problem-solving skills Judgement based on experience

Appendix B

Outdoor Situational Leadership Rubric

Outdoor Situational Leadership Rubric

		0	1	2	3
		no indicators present	indicator undefined, ambiguous or unexplored	indicator stated, described some aspects lack clarity.	indicator explained clearly and described comprehensively, delivering all relevant information necessary for full understanding.
Ability See journal part e	Knowledge				
	Skill				
	Experience				
Willingness See journal part e	Confidence				
	Commitment				
	Motivation				
Performance Readiness Application See journal parts d and e		0 mismatch or unable to assess	3 match		
			Selects the correct style for the readiness level (compare part b & d) and...		
Leadership Application See journal part c		0 Does not apply the appropriate leadership style for the performance readiness of the group. Style is not correct for readiness level (compare part b & d)	1 the description of the style is missing, vague, or inconsistent with the leadership style	2 the description of the style is described, some aspects lack clarity.	3 provides an explanation of the style stated clearly and described comprehensively, delivering all relevant information necessary for full understanding using verbs (eg. coaching, directing, supporting) which are consistent with the style according to the SLM

Appendix C

Structured Journal

Please create a confidential identifier as follows. Your first initial, birth month (spelled out), and last two digits of your phone number. For example, GJuly22 Identifier _____

Instructions

1. You will complete your journal when you are leader of the day. The journal entry should be completed in the evening during the designated journaling time.
2. Record one entry for each of the **three** most important **leader** decisions of the day.
    a. Leader decisions affect someone else or the group. Leader decisions are not personal decisions, like deciding when to put on an extra layer.
3. The journal entry should reflect on your own decision-making. Not a critique of someone else's decision-making.

Date_____ Decision Number_____

a. Describe the situation (what was the task?).

b. What style did you use to make this decision? Circle ONE:

S1 (Telling) S2 (Selling) S3 (Participating) S4 (Delegating)

c. Describe how you implemented the selected style using verb indicators.

d. Select the **readiness** level of the group for this decision. Circle ONE:

R4 (Able/Willing and confident) R3 (Able/Unwilling or insecure) R2 (Unable/Willing or Confident) R1 (Unable/Unwilling or insecure)

e. Describe the behaviors you observed (indicators) of ability (knowledge, skill, experience) and willingness (confidence, commitment, motivation) the group demonstrated that led you to choose the readiness level.

References

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Asfeldt, M., & Hvenegaard, G. (2014) Perceived learning, critical elements, and lasting impacts on university-based wilderness educational expeditions. *Journal of Adventure Education and Outdoor Learning*, *14*(2), 132-152. https://doi.org/10.1080/14729679.2013.789350

Astin, A. W. (1993). What matters in college? Four critical years revisited (Vol. 1). San Francisco: Jossey-Bass.

Attarian, A. (2001). Trends in outdoor adventure education. *Journal of Experiential Education, 24*(3), 141–149. https://doi.org/10.1177/105)382590102400304

Anderson, D. W., Krajewski, H.T., Goffin, R.. D., & Jackson, D. N. (2008). A leadership self-efficacy taxonomy and its relation to effective leadership, *The Leadership Quarterly,19(5),* 595-608. https://doi.org/10.1016/j.leaqua.2008.07.003

Andre, E. K., Williams, N., Schwartz, F., Bullard, C. (2017). Benefits of campus outdoor recreation programs: A review of the literature. *Journal of Outdoor Recreation*, Education, *and Leadership*, *9,* 15–25. https://doi.org/10.18666/JOREL-2017-V9-I1-7491

Antonakis, J., Avolio, B., & Sivasubramaniam, N. (2003). Context and leadership: an examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire, *The Leadership Quarterly,*14 (3), 261-295. https://doi.org/10.1016/S1048-9843(03)00030-4

Bachert, D. (1990). Historical evolution of NOLS: The National Outdoor Leadership School. In J.C. Miles & S. Priest (Eds.), *Adventure Education* (pp. 83-88). Venture.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-212.

Bass, B. M. (1999). Two decades of research and development in transformational leadership. *European Journal of Work and Organizational Psychology*, 8(1), 9-32.

Bandalos, D.L. (2018). *Measurement theory and applications for the social sciences.* New York. Guilford Press.

Bell, B. J., Holmes, M. R., & Williams, B. G. (2010). A census of outdoor orientation programs at four-year colleges in the United States. *Journal of Experiential Education, 33*(1), 1–18. https://doi.org/10.1177/105382591003300102

Bell, B. J., Gass, M. A., Nafziger, C. S., & Starbuck, J. D. (2014). The state of knowledge of outdoor orientation programs: Current practices, research, and theory. *Journal of Experiential Education*, *37*(1), 31–45. https://doi.org/10.1177/1053825913518891

Berman, D., & Teeters, C. (2003). Wilderness Education Association Affiliate Handbook.

Blanchard, J., Strong, M., & Ford, P. M. (2007). *Leadership and administration of outdoor pursuits*. Venture Pub.

Blumenstyk, G. (2012, January 30). College officials welcome Obama's focus on higher-education costs, but raise some concerns. The Chronicle of Higher Education. Retrieved from http://chronicle.com/article/President-Puts-College-Costs/130503/

Bobilya, A. J., Holman, T., Lindley, B., Faircloth B. W. & Homan, T. (2017). A qualitative analysis of participant learning and growth using a new Outward Bound outcomes instrument. J*ournal of Outdoor Recreation, Education, and Leadership, 9*(2), 254-257.

Brown, M. (2004). “Let’s go round the circle”: How verbal facilitation can function as a means of direct instruction. *Journal of Experiential Education, 27*(2), 161-175.

Brown, M. (2009). Reconceptualizing outdoor adventure education: Activity in search of an appropriate theory. *Australian Journal of Outdoor Education, 13*(2), 3-13.

Brymer, E. & Gray, T. (2006). Effective leadership: Transformational or transactional? *Australian Journal of Outdoor Education, 10*(2), 13-19.

Brymer, E., Gray, T., Carpenter, C., & Cotton, W. (2010). Profiling outdoor leadership. *Journal of Outdoor Recreation, Education, and Leadership, 2(1),* 93-108.

Buell, L.H. (1981). *The identification of outdoor adventure leadership competencies for entry-level and experience-level personal.* (Unpublished doctoral dissertation). University of Massachusetts, Amherst, MA.

Burns, K.J. (1998), Classical reliability: Using generalizability theory to assess dependability. *Research in Nursing & Health*, 21, 83-90. https://doi.org/10.1002/(SICI)1098-240X(199802)21:1<83::AID-NUR9>3.0.CO;2-P

Cain, K., & McAvoy, L. (1990). Experience-based judgment. In J. Miles & S. Priest (Eds.), *Adventure education* (pp. 241-250). State College, PA: Venture Publishing.

Cockrell, D. (1991). *The Wilderness Educator: The Wilderness Education Association Curriculum Guide*. ICS Books, Inc.

Coley, S. J., Burns, V. E., & Cumming, J. (2015). The role of outdoor adventure education in facilitating groupwork in higher education. *Higher Education*, *69*(4), 567–582. https://doi.org/10.1007/s10734-014-9791-4

Crawford, J. A., & Kelder, J. (2019). Do we measure leadership effectively? Articulating and evaluating scale development psychometrics for best practice. *The Leadership Quarterly*, *30*(1), 133-144. https://doi.org/10.1016/j.leaqua.2018.07.001

Day, D., & Antonakis, J. (2012). Leadership: Past, present, and future. In Antonakis, J. and Day, D. (Eds.), *The nature of leadership.* Sage Publications, Inc.

Drury, J. K., Bonney, B. F., Berman, D., & Wagstaff, M.C. (1992). *The Backcountry Classroom: Lesson Plans for Teaching in the Wilderness*. Globe Pequot Press

Ewert, A., & Overholt, J. (2010). Fostering leadership through a three-week experience: Does outdoor education make a difference? *Research in Outdoor Education*, *10*, 38–47. https://doi.org/10.1353/roe.2010.0005

Fernandez, C. F., & Vecchio, R. P. (1997). Situational leadership theory revisited: A test of an across–jobs perspective. *The Leadership Quarterly*, *8*(1), 67–84. https://doi.org/10.1016/S1048-9843(97)90031-X

Ford, P., & Blanchard, J. (1985). *Leadership and administration of outdoor pursuits*. State College, PA: Venture Publishing.

Frauman, E., & Anderson, C. (2018, October). *Examining perceived confidence and proficiency of college students participating in an outdoor leadership development course: An exploratory study.* Poster session presented at the Association of Outdoor Recreation and Education Annual Conference, Snowbird, UT.

Friel, B. (2003). The ivory climbing wall. National Journal, 35, 3488–3494

Gabriel, J. (2015). *Situational leadership awareness development in student outdoor leaders through training versus experience*. [Doctoral dissertation, Bowling Green State University]. ProQuest Dissertations & Theses Global.

Galloway, S. P. (2000). Assessment in wilderness orientation programs: Efforts to improve college student retention. *Journal of Experiential Education, 23*(2), 75–84. https://doi.org/10.1177/105382590002300205

Gass, M. A. (1999). Accreditation and certification: Questions for an advancing profession. In Miles, J. C., & Priest, S. (Eds.). *Adventure Programming* (pp. 247-252). State College, PA: Venture.

Gookin, J., & Leach, S. (Eds.). (2009). *2009 NOLS Leadership Educator Notebook: A Toolbox for Leadership Educators*. National Outdoor Leadership School.

Graham, J. (1997). *Outdoor leadership: Technique, common sense and self-confidence*. Seattle, WA: The Mountaineers.

Green, P.J. (1981). *The content of a college-level outdoor leadership course for land-based outdoor pursuits in the Pacific Northwest: A Delphi consensus.* (Unpublished doctoral dissertation). University of Oregon, Eugene, OR.

Grube, D. P., Phipps, M. L., & Grube, A. J. (2002). Practicing leader decision-making through a systematic journal technique: A single case analysis. *Journal of Experiential Education, 25*(1), 220–230. https://doi.org/10.1177/105382590202500106

Hannah, S. T., Avolio, B. J., Luthans, F., and Harms, P. D. (2008). Leadership efficacy: Review and future directions, *The Leadership Quarterly, 19(6),* 669-692. https://doi.org/10.1016/j.leaqua.2008.09.007

Hattie, J., Marsh, H. W., Neill, J. T., & Richards, G. E. (1997). Adventure education and Outward Bound: Out-of-class experiences that make a lasting difference. *Review of Educational Research, 67*(1), 43–67.

Huey, W. S., Smith, D. G., Thomas, J. J., & Carlson, C. R. (2014). The great outdoors: Comparing leader development programs at the U.S. Naval Academy. *Journal of Experiential Education, 37,* 367-381. https://doi.org/10.1177/1053825914547625

Harmon, B. V. 2006. A qualitative study of the learning processes and outcomes associated with students who serve as peer mentors. *Journal of The First-Year Experience and Students in Transition, 18*(2), 53–82.

Hayashi, A., & Ewert, A. (2006). Outdoor leaders' emotional intelligence and transformational leadership. *Journal of Experiential Education, 28*(3), 222–242. https://doi.org/10.1177/105382590602800305

Hersey, P., Blanchard, K., & Johnson D. (2013). Management of organizational behavior: Utilizing human resources. (10th. ed.). Prentice Hall.

Hubbs, & Brand, C. F. (2005). The Paper Mirror: Understanding Reflective Journaling. *The Journal of Experiential Education, 28*(1), 60–71. https://doi.org/10.1177/105382590502800107

Johansen, B. C. P. (1990). Situational leadership: A review of the research. *Human Resource Development Quarterly, 1*(1), 73-85.

Johnson, R. L., Penny, J., Gordon, B., Shumate S. R., & Fisher, S. P. (2005) Resolving score differences in the rating of writing samples: Does discussion improve the accuracy of scores?, *Language Assessment Quarterly, 2*(2), 117-146, https://doi.org/10.1207/s15434311laq0202_2

Jordan, D. (1989) A new vison for outdoor leadership theory. *Leisure Studies: The Journal of the Leisure Studies Association, 8*(1), 35-47.

Kass, D., & Grandzol, C. (2012). Evaluating the value-added impact of outdoor Management training for leadership development in an MBA program. *Journal of Experiential Education*, *35*(3), 429-446.

Kosseff, A. (2003). *AMC guide to outdoor leadership*. Boston, MA: Appalachian Mountain Club.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21. https://doi.org/10.3102/0013189X020008015

Lord, R., & Dinh, J. (2012). Aggregation processes and levels of analysis as organizing structures for leadership theory. In Antonakis, J. and Day, D. (Eds.), *The nature of leadership.* Sage Publications, Inc.

Lupton, F. (1990). WEA history. In J.C. Miles & S. Priest (Eds.), *Adventure Education* (pp. 89-95). Venture.

Martin, A. (2012, December 13). Building a showcase campus, using an I.O.U. *New York Times*. Retrieved from http://www.nytimes.com/2012/12/14/business/colleges-debt-falls-onstudents-after-construction-binges.html?pagewanted=all

Martin, B., Cashel, C., Wagstaff, M., & Breunig, M. (2006). *Outdoor leadership: Theory and practice*. Human Kinetics.

McKenzie, M. (2003). Beyond "the outward B\bound process": Rethinking student learning. *Journal of Experiential Education*, 26(1), 8–23 https://doi.org/10.1177/105382590302600104

Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher, 18(2),* 5–11. https://doi.org/10.2307/1175249

Messick, S. (1995). Standards of validity and the validity of standards in performance asessment. *Educational measurement: Issues and practice*, *14*(4), 5-8.

Miles, J. C., & Priest, S. (1990). *Adventure education*. Venture Publishing, Inc.

Moskal, B.M. and Leydens, Jon A. (2000). Scoring rubric development: Validity and reliability, *Practical Assessment, Research, and Evaluation*, 7(10), DOI: https://doi.org/10.7275/q7rm-gg74

NIRSA (2021). *NIRSA institutional data set.* NIRSA. https://nirsa.net/nirsa/research-assessment/institutional-data-set/

Ogilvie, K. C. (2005). *Leading and managing groups in the outdoors* (2nd ed.). Penrith: The Institute for Outdoor Learning.

Outward Bound International (n.d.). https://www.outwardbound.org/about-us/

Paisley, K., Furman, N., Sibthorp, J., & Gookin, J. (2008). Student learning in outdoor education: A case study from the National Outdoor Leadership School. *Journal of Experiential Education*, *30*(3), 201–222. https://doi.org/10.5193/jee.30.3.201

Peart, R. (1991). Safe Leadership in OE: Part 2: Effective Leadership. *Journal of Adventure Education and Outdoor Leadership*, 8(4), 21-25.

Pelchat, C. & Karp, G.G. (2012). A historical review of outdoor leadership, curricular development, and the future with action research. *SchOSLR: A Journal of Leisure Studies and Recreation Education*, 27 (2), 1-13.

Petzoldt, P. (1984). *The new wilderness handbook.* New York: W. W. Norton and Company.

Pierce, J. L., & Newstrom, J. W. (2003). Leaders and the leadership process: Readings, self-assessments, and application (3rd ed.). New York: McGraw.

Phipps, M. L. (1986). An assessment of a systematic approach to teaching outdoor leadership in expedition settings. Unpublished doctoral dissertation, University of Minnesota.

Phipps, M. L. (1987). Group dynamics in the outdoors: A model for teaching outdoor leaders. High Adventure Outdoor Pursuits: Organization and leadership. Meier, J., Morash, T., Welton, G. (Eds.) Publishing Horizons, Inc.

Phipps, M. L. (1996). A systematic approach to learning leadership. Unpublished course package, Western Carolina University, Cullowhee, NC.

Phipps, M. L. (1992). The ELSA manual. Sylva, NC: System 4 Services.

Phipps, M. L., Hayashi, A., Lewandowski A., & Padgett, A. H. (2005) Teaching and evaluating instructor effectiveness using the instructor effectiveness questionnaire and the instructor effectiveness check sheet combination. *Journal of Adventure Education and Outdoor Learning, 5*(1), 51-65, https://doi.org/10.1080/14729670585200591

Phipps, M., & Swiderski, M. (1990). The "soft" skills of outdoor leadership. *Adventure education*, 221-232.

Phipps, M. L., & Phipps, C. A. (2003). Using the Expedition Leader Style Analysis.

Popham, J. W. (1997). What's wrong—and what's right—with rubrics. *Educational Leadership, 55*(2), 72–75.

Povilaitis, V., Riley, M., DeLange, R., Verkouw, AJ., Macklin, K., & Hodge, C. J. (2019). Instructor impacts on outdoor education participant outcomes: A systematic review. *Journal of Outdoor Recreation, Education, and Leadership, 11*(3), 222-238.

Priest, S. (1984). Effective outdoor leadership: A survey. *The Journal of Experiential Education*, 7(3), 34-36.

Priest, S. (1986). *Outdoor leadership in five nations* (Unpublished doctoral dissertation). University of Oregon, Eugene, OR.

Priest, S. (1988). Agreement reached on the issue of outdoor leadership certification? *Bradford Papers Annual, 3,* 38-43.

Priest, S. (1987). *Preparing effective outdoor pursuit leaders.* Institute of Recreation Research and Service. Eugene, OR.

Priest, S., & Chase (1989). The conditional theory of outdoor leadership: An exercise in flexibility. *Journal of Adventure Education and Outdoor Leadership, 6(2),* 10-17.

Priest, S., & Dixon, T. (1991). Research note: towards a new theory of outdoor leadership style. *Leisure studies: The Journal of the Leisure Studies Association*, *10*(1), 163-170.

Priest, & Gas, M. A. (1997). *Effective leadership in adventure programming (1st ed.).* Champaign, IL: Human Kinetics.

Priest, & Gas, M. A. (2018). *Effective leadership in adventure programming (3rd ed.).* Champaign, IL: Human Kinetics.

Propst, D. B., & Koesler, R. A. (1998) Bandura goes outdoors: ROSLR of self-efficacy in the outdoor leadership development process, *Leisure Sciences*, *20*(4), 319-344. https://doi.org/10.1080/01490409809512289

Raiola, E.O. (1986). *Outdoor wilderness education: A leadership curriculum.* (Unpublished doctoral dissertation). Union Graduate School, Potsdam, NY.

Richards, A. (1990). Kurt Hahn. In J.C. Miles & S. Priest (Eds.), *Adventure Education* (pp. 67-74). Venture.

Rilling, C., & Jordan, D. (2007). Important co-leader skills and traits on extended outdoor trips as perceived by leaders, *Leisure Studies, 26*(2), 193-212. https://doi.org/10.1080/02614360500337524

Russel, J. H., & Skinkle, R., R. (1990). Evaluation of peer-adviser effectiveness. *Journal of College Student Development, 31*(5), 388–394.

Sawaki, Y. (2012). Generalizability theory. In *Encyclopedia of Research Design* (pp. 534-538). Thousand Oaks: Sage.

Seaman, J., Bell, B. J., & Trauntvein, N. (2017). Assessing the value of a college degree in outdoor education or recreation: Institutional comparisons using the College Scorecard and surveys of faculty and employers. *Journal of Outdoor Recreation, Education, and Leadership, 9*(1), 26-41.

Scott, A. (2012, July 26). Climbing walls and college costs. Retrieved from http://www.marketplace.org/2012/07/26/education/climbing-walls-and-college-costs

Smith, H., & Penny, D. (2010). Effective, exemplary, extraordinary? Towards an understanding of extraordinary outdoor leadership. *Australian Journal of Outdoor Education, 14*(1), 23-29.

Speelman, E.A. & Wagstaff, M. (2015), Adventure leadership and experiential education. *New Directions for Student Leadership*, 147, 89-98. https://doi.org/10.1002/yd.20146

Starbuck, D. J., & Bell, B. J. (2017). Outdoor orientation leaders: The effects of peer leadership. *Journal of Outdoor Recreation Education and Leadership*, 9(3), 280-297.

Stremba, B., & Bisson, C. A. (Eds.). (2009). *Teaching adventure education theory: Best Practices.* Champaign, IL. Human Kinetics.

Swiderski, M.J. (1981). *Outdoor leadership competencies identified by outdoor leaders in five western regions* (Unpublished doctoral dissertation). University of Oregon, Eugene, OR.

Tannenbaum, & Schmidt, W. H. (1986). How to choose a leadership pattern. *Harvard Business Review*, *64*, 129.

Thompson, G., & Glasø, L. (2018). Situational leadership theory: a test from a leader-follower congruence approach. *Leadership & Organization Development Journal*, *39*(5), 574-591.

Thompson, G., & Vecchio, R. P. (2009). Situational leadership theory: A test of three versions. *The Leadership Quarterly, 20*(5), 837–848. https://doi.org/10.1016/j.leaqua.2009.06.014

Vecchio, R.B. & Boatwright, K. J. (2002). Preferences for idealized styles of supervision, *The Leadership Quarterly, 13(4),* 327-342.

Vecchio, R. P., Bullis, R. C., & Brazil, D. M. (2006). The Utility of Situational Leadership Theory: A Replication in a Military Setting. *Small Group Research*, *37*(5), 407–424. https://doi.org/10.1177/1046496406291560

Warren, K. (2002). Preparing the next generation: Social justice in outdoor leadership education and training. *Journal of Experiential Education*, *25*(1), 231–238. https://doi.org/10.1177/105382590202500107

Webb, D. J. (2001). The emergence and evolution of outdoor adventure programs, 1863-2000: A history of student initiated outing programs. *Proceedings of the Annual International Conference on Outdoor Education*. https://files.eric.ed.gov/fulltext/ED476477.pdf

Wilderness Education Association (n.d.). https://www.weainfo.org/oltc

Woodhouse, K. (2015, June 15). Lazy rivers and student debt. *Inside Higher Ed.* Retrieved from https://www.insidehighered.com/news/2015/06/15/are-lazy-rivers-and-climbing-wallsdriving-cost-college