

## Research Note

# Is “Acceptable” Really Acceptable? Comparing Two Scales for Measuring Normative Evaluations in Outdoor Recreation

ROBYN L. CEURVORST

Department of Environment and Society  
Utah State University  
Moab, UT, USA

MARK D. NEEDHAM

Department of Forest Ecosystems and Society  
Oregon State University  
Corvallis, OR, USA

*This research note compares the commonly used scale of “very unacceptable” to “very acceptable” with an alternative scale of “should definitely not allow” to “should definitely allow” for measuring normative evaluations. Data were obtained from 1,422 users at six sites in Hawaii, and norms were measured in questionnaires with photographs depicting increasing use densities. Responses differed statistically between scales with users feeling that there should be more people allowed than they thought were acceptable at each site. The “acceptance” scale also generated more crystallization or agreement, but lower intensity or importance of use levels than the scale based on “should.” Effect sizes, however, showed that the magnitude of these differences was small or minimal, suggesting that the “acceptance” scale may be appropriate.*

**Keywords** crystallization, encounters, intensity, norms

## Introduction

The concept of “norms” has received considerable attention in the recreation and leisure literature (see Manning, 2007, 2011; Shelby, Vaske, & Donnelly, 1996; Vaske & Whittaker, 2004, for reviews). One line of research defines norms as what people think social, resource,

Received 28 January 2011; accepted 28 January 2012.

Address correspondence to Mark D. Needham, Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331. E-mail: mark.needham@oregonstate.edu

Support was from the Hawaii Coral Reef Initiative Research Program and Hawaii Division of Aquatic Resources pursuant to National Oceanic and Atmospheric Administration awards NA06NOS4260200 and NA05NOS4191014. The editors, associate editors, and reviewers are thanked for comments on earlier versions. The lead author was involved in this research as part of her doctoral program at Oregon State University.

or facility conditions should or should not be and are judgments for evaluating activities, environments, or management strategies as good or bad, better or worse (Shelby et al., 1996; Vaske, Shelby, Graefe, & Heberlein, 1986). Researchers have used response scales such as “undesirable” to “desirable,” “unpleasant” to “pleasant,” and “not preferable” to “preferable” in questionnaires measuring norms (see Manning, Valliere, Wang, & Jacobi, 1999, for a review). The most frequently used scale has been “unacceptable” to “acceptable” (see Manning, 2011; Shelby et al., 1996, for reviews). Research has shown, however, that respondents can attach different meanings to these scales. Users, for example, often prefer seeing fewer than half as many people as their maximum tolerance limit (Manning, 2007; Manning et al., 1999).

Some researchers have suggested that evaluations such as “preference” and “acceptance” may not be consistent with many conventional definitions of norms or directly measure the concept because they may not imply an inherent sanction or impose a sense of obligation to conform to a norm (e.g., Heywood, 1996a, 1996b, 2002, 2011; Roggenbuck, Williams, Bange, & Dean, 1991). It has been suggested that these evaluations arise from social regularities and may not be purely normative, but more along the lines of social conventions, attitudes, or emerging norms (Heywood, 1996a, 2011). Social regularities and conventions may become norms as individuals recognize implied obligations in reference to particular social groups and situations (Cancian, 1975). If everyone in a group supports the norm, it is in each individual’s interest that it be followed to avoid experiencing sanctions (e.g., embarrassment, dirty looks; Opp, 1982). Several researchers, therefore, have suggested that measurement of norms must at least use terms that are arguably more unambiguous and straightforward, such as “should” and “ought,” because they conform to most definitions of norms and may convey a clearer sense of obligation (Blake & Davis, 1964; Cancian, 1975; Heywood, 1996a, 2002; Opp, 1982).

Researchers have suggested that more studies are needed examining these evaluative dimensions and if they differentially influence normative responses (e.g., Heywood, 1996a, 2002; Manning et al., 1999). This research note, therefore, is methodological and empirically compares the commonly used scale of “very unacceptable” to “very acceptable” with an alternative scale of “should definitely not allow” to “should definitely allow.” Three research questions are examined. First, what are normative evaluations of users when measured with “acceptance” of use densities? Second, what are normative evaluations of users when measured based on the term “should?” Third, do evaluations differ between these two scales?

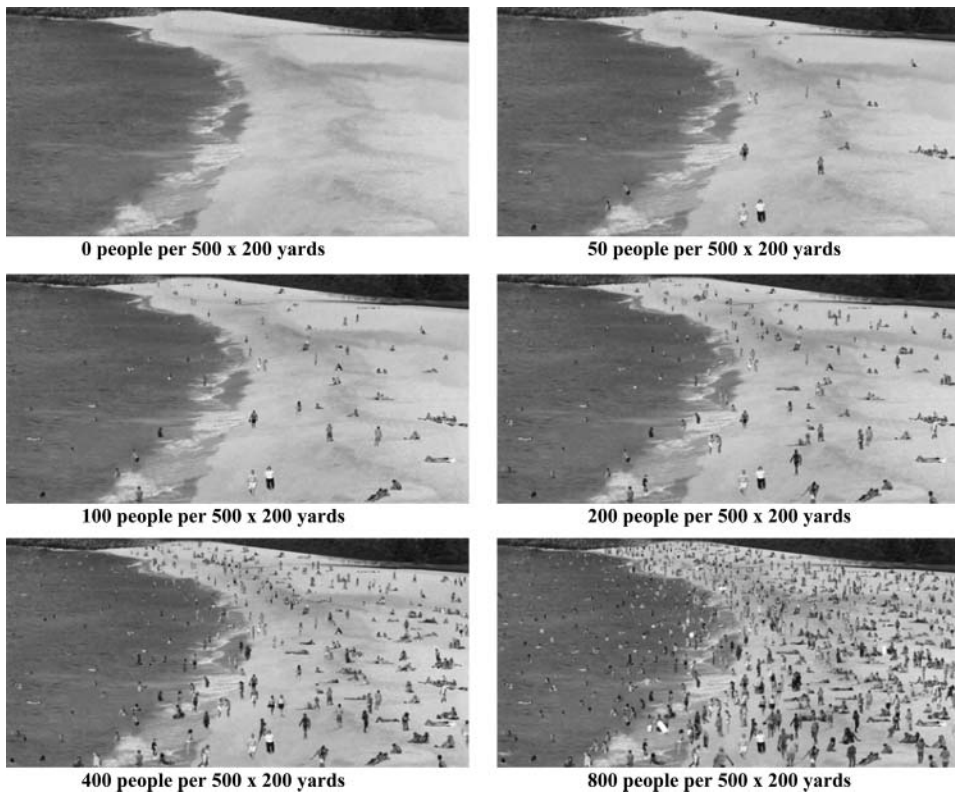
## Methods

Data were obtained in July and August 2007 from a survey of summer recreationists at coastal sites on the island of Oahu, Hawaii. These sites were Kailua Beach Park, two sites at Waikiki-Diamond Head Shoreline Fisheries Management Area (Sans Souci/Kaimana Beach, Diamond Head Beach Park), and three sites at Pupukea Marine Life Conservation District (Waimea Bay, Three Tables, Shark’s Cove). There is marginal seasonal variation in visitation to these sites and although they have regulatory and jurisdictional differences ranging from a county park to a state marine protected area, they have similar amenities, activities, and resources (Needham et al., 2008). Individuals completed questionnaires onsite at the beach, shore, and park areas. Questionnaires were administered at each site at least once for each day of the week and for each of three time periods (8:00–10:30 a.m., 11:30 a.m.–2:00 p.m., and 3:00–5:30 p.m.). Individuals were selected through systematic random sampling (one random person from every 5th or 10th group, depending on the site), and 1,422 questionnaires were completed with an 87% response rate (Kailua Beach: 476,

85%; Waikiki–Diamond Head: 462, 84% [Sans Souci/Kaimana: 289, 90%; Diamond Head Beach: 173, 75%]; Pupukea: 484, 93% [Waimea Bay: 198, 94%; Three Tables: 145, 92%; Shark’s Cove: 141, 93%]).

Consistent with recent studies, photographs were used for depicting use densities (see Manning, 2007; Manning & Freimund, 2004, for reviews). It has been suggested that visuals can be more beneficial than written approaches for measuring norms because they allow respondents to see conditions and researchers to depict a range of impacts (Hall & Roggenbuck, 2002; Manning & Freimund, 2004). Respondents viewed six color photographs of 0–800 people per 500 × 200 yards with the number doubling in each image (0, 50, 100, 200, 400, 800 people; Figure 1). To reflect site use patterns on most days, densities were divided with 70% of people in each image on land (beach, shore) and 30% in the ocean. Approximately half of the image width was land (100 yards) and half was ocean (100 yards); the length was the same for both (500 yards). Using Adobe Photoshop, the image of 800 people was created first, and people were randomly removed to create five other photographs. People were randomly positioned, but their age, gender, and number in the foreground and background were balanced. The density scale was measured in the field at 500 × 200 yards. Similar to other studies, respondents were told to ignore the generic backgrounds in the visuals, focus on use densities, and assume that they were occurring at the site where they were surveyed (e.g., Manning, 2007; Needham et al., 2011).

Respondents rated conditions in each image first on the common 9-point scale of “very unacceptable” to “very acceptable.” Following this, a second set of questions used an alternative 9-point scale of “should definitely not allow” to “should definitely allow”



**FIGURE 1** Use density photographs.

with interior narratives of “should maybe not allow” and “should maybe allow.” This scale eliminated any temporal components inferred in scales with similar wording (e.g., “should never,” “should always”; Heywood, 1996a, 2002). Debriefing with respondents during questionnaire pretests suggested that they had no difficulty interpreting these scales and photographs.

## Results

Responses from these scales are provided in Table 1 and shown as mean norm curves in Figure 2. The minimum acceptable condition is where the curve crosses the neutral line and conditions become unacceptable or intolerable for the majority of respondents (Shelby et al., 1996). These conditions ranged from 192 to 381 people per 500 × 200 yards, and were lower across all six sites on the “acceptance” scale compared to the scale based on “should.” This condition at Kailua Beach Park, for example, was 302 people per 500 × 200 yards using “acceptance,” whereas it was 340 people on the “should” scale. Paired-sample *t*-tests were statistically significant at five sites and across all sites combined,  $t = 2.74$  to  $9.44$ ,  $p = .007$  to  $< .001$ . The exception was Diamond Head Beach Park where this pattern of

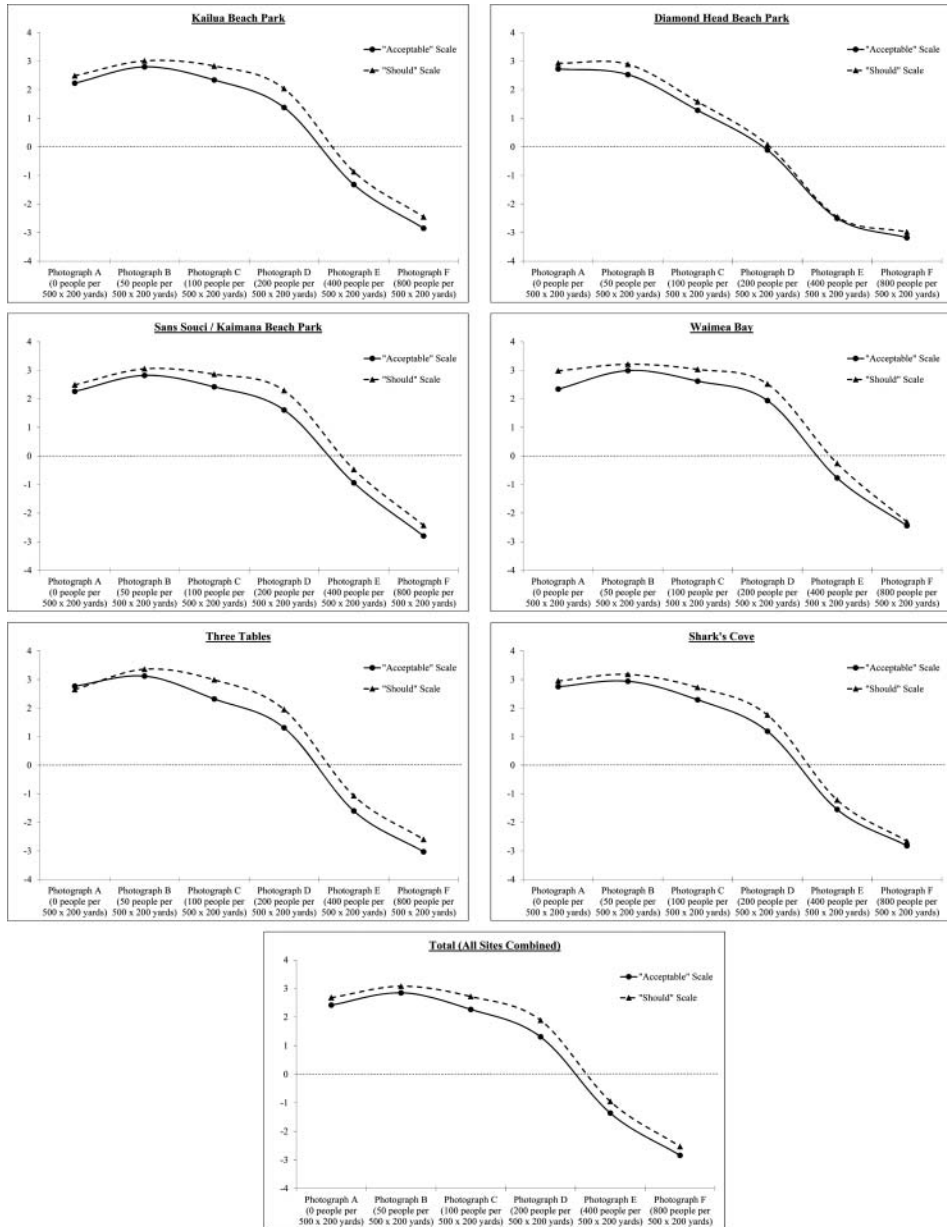
**TABLE 1** Normative Evaluations for Each Scale at Each Site

	“Acceptable” Scale	“Should” Scale	Paired-sample <i>t</i> -test value	<i>p</i> -value	Cohen’s <i>d</i> effect size
Minimum acceptable condition <sup>1</sup>					
Kailua Beach	302.22	340.21	6.31	<.001	.05
Diamond Head Beach	192.09	206.32	0.77	.442	.01
Sans Souci/Kaimana	326.27	365.94	3.93	<.001	.05
Waimea Bay	343.17	381.29	3.96	<.001	.08
Three Tables	290.03	329.14	3.95	<.001	.06
Shark’s Cove	286.86	318.12	2.74	.007	.04
Total	298.13	333.10	9.44	<.001	.04
Intensity (maximum = 24) <sup>2</sup>					
Kailua Beach	12.92	13.69	4.88	<.001	.06
Diamond Head Beach	12.33	12.91	1.23	.324	.04
Sans Souci/Kaimana	12.85	13.59	3.15	.002	.06
Waimea Bay	13.10	14.31	3.32	.001	.10
Three Tables	14.15	14.62	2.92	.004	.04
Shark’s Cove	13.54	14.49	2.28	.024	.08
Total	13.05	13.85	7.31	<.001	.06
Crystallization <sup>3</sup>					
Kailua Beach	2.01	2.14	12.55	<.001	—
Diamond Head Beach	2.03	2.17	4.94	<.001	—
Sans Souci/Kaimana	2.05	2.15	4.59	<.001	—
Waimea Bay	2.06	2.06	1.06	.638	—
Three Tables	1.83	2.10	3.11	<.001	—
Shark’s Cove	1.84	1.97	3.61	<.001	—
Total	2.02	2.16	22.99	<.001	—

<sup>1</sup>Mean number of people where curve crosses neutral line.

<sup>2</sup>Distance from neutral line across all curve points independent of evaluation direction.

<sup>3</sup>Average standard deviations across all curve points. Test values represent Levene’s test for homogeneity.



**FIGURE 2** Norm curves for each site on 9-point scales of -4 “should definitely not allow” or “very unacceptable” to +4 “should definitely allow” or “very acceptable.”

differences between scales was evident (“accept” = 192, “should” = 206) but insignificant at  $t = 0.77$ ,  $p = .442$ . On average, users felt that there should be more people allowed than they thought were acceptable at each site. Cohen’s  $d$  effect size statistics, however, ranged from only .01 to .08. Using guidelines from Cohen (1988) and Vaske (2008) for interpreting effect sizes, these results suggest that although there were statistical differences between scales, the magnitude of these differences was “small” or “minimal,” respectively.

The slope of the curve and cumulative summed distance from the neutral line at each point irrespective of direction (negative, positive) is one measure of norm intensity or salience, or the importance of use densities to respondents (Needham et al., 2011; Vaske et al., 1986). At all sites, this value was higher on the scale based on “should.” At Kailua Beach Park, for example, this value was higher on the “should” scale (13.69) than “acceptance” (12.92). This pattern was statistically significant at five sites and across all sites combined,  $t = 2.28$  to  $7.31$ ,  $p = .024$  to  $< .001$ . Again, the exception was Diamond Head Beach Park where this pattern of differences was evident (“accept” = 12.33, “should” = 12.91) but insignificant,  $t = 1.23$ ,  $p = .324$ . Use densities, therefore, were rated as more important on the “should” scale. Cohen’s  $d$  effect sizes, however, ranged from only .04 to .10, suggesting that these differences were “small” (Cohen, 1988) or “minimal” (Vaske, 2008).

Crystallization indicates potential agreement among respondents, and the average of the standard deviations across all points on the curve is one measure of crystallization (Needham et al., 2011; Shelby et al., 1996). At most sites, there was more crystallization (i.e., lower standard deviations) when evaluations were based on “acceptance.” At Three Tables, for example, there was more crystallization ( $SD = 1.83$ ) using “acceptance” than the “should” scale ( $SD = 2.10$ ). Levene’s tests for homogeneity revealed that differences between scales were significant at five sites and for all sites combined, Levene’s  $F = 3.11$  to  $22.99$ ,  $p < .001$ . At Waimea Bay, however, crystallization was identical for each scale ( $SD = 2.06$ ).

## Discussion

This research note compared the commonly used scale of “very unacceptable” to “very acceptable” with an alternative scale of “should definitely not allow” to “should definitely allow” for measuring norms at coastal sites in Hawaii. Responses to these scales differed, with users feeling that there should be more people allowed than they thought were acceptable at each site. The “acceptance” scale also generated more crystallization or agreement, but lower intensity or importance of use levels than the scale based on “should.” Although these differences were statistically significant in most cases, effect sizes suggested that the magnitude of these differences was small or minimal. Regardless, managers should be mindful of differences among evaluations when establishing and monitoring indicators and standards of quality. If standards are based on the “should” scale, for example, conditions may still be unacceptable for the majority of visitors. Researchers should also clarify the meanings of scales when reporting findings to managers.

These differences based on scale terminology are somewhat consistent with past research. Studies have found, for example, that people often “prefer” fewer encounters than they “accept,” but their “maximum tolerance” is often higher than both preference and acceptance (Manning et al., 1999). Research has also shown that users often “accept” fewer encounters than they will tolerate before “management action” is deemed necessary (Manning et al., 1999). Similarly, results here revealed that users thought that more people should be allowed at each site than they would accept. It is possible, therefore, that some people may interpret the “management action” response used in previous studies somewhat similarly to the scale based on “should” used here. This suggests that conditions that should not be allowed or have deteriorated to a point where management attention is deemed necessary may imply an obligation or institutional sanction.

These differences suggest that users attach different meanings to scales. It is also possible that evaluations based on terms such as “accept” and “should” measure different dimensions of norms such as personal, social, and institutional norms (Heywood, 2011; Vaske & Whittaker, 2004). These scales may also be measuring different things, as some

researchers have claimed that evaluations such as “acceptance” and “preference” are social conventions or cognitively generated evaluative standards about conditions such as better or worse use densities, and these are not in line with conventional definitions of norms (Heywood, 1996a, 2011; Roggenbuck et al., 1991). It has been suggested that norms are attached to terms such as “should” and “ought,” which are more likely to be interpreted as implying more informed normative judgments than other scales (Blake & Davis, 1964; Cancian, 1975; Heywood, 2002). These terms may also imply stronger social obligations to conform or comply with behaviors and conditions such as a sense of obligation about personal space and territory (Heywood, 1996b, 2011; Opp, 1982).

The title of this research note asks if the commonly used “acceptance” scale is acceptable, but the goal is not to argue that one evaluative term is better than another. Responses statistically differed between the scales, but effect sizes showed that these differences were not substantial. Across all sites combined, for example, users “accepted” no more than 298 people per 500 × 200 yards but felt that no more than 333 people “should” be allowed—a difference of only 35 people in an area the size of approximately five American football fields. The “acceptance” scale used in most recreation and leisure studies, therefore, may be appropriate, but it remains a question of future research to determine if responses based on terms such as “accept” or “should” actually imply any normative sanctions or generate a clear sense of obligation among users.

Respondents evaluated visuals using the “acceptance” scale first and then the scale based on “should.” Evaluations may have been influenced by this ordering of questionnaire variables. However, studies have examined issues such as starting point bias by comparing how ordering of questions and visuals influences user norms, and results have showed minimal differences (Manning et al., 1999; Needham et al., 2011). Given that ordering effects were not tested here, research should examine this by possibly using an experimental approach. Two questionnaire versions, for example, could be administered with one using “acceptance” first and the other with the “should” scale first. Statistical tests could examine possible response differences. The visual approach used here is also consistent with recent recreation and leisure studies (see Manning, 2007; Manning & Freimund, 2004, for reviews), but research is needed using measures of “should” in traditional narrative and closed-ended questions or open-ended and fill-in-the-blank written formats (Shelby et al., 1996). This could allow comparisons with other normative evaluations (e.g., “acceptance,” “preference”) using these written formats to determine if results generalize across methodological approaches. Studies should also examine differences in scales for indicators other than encounters and use densities (e.g., litter, bare ground) because one approach may be more appropriate than another depending on the indicators examined. Finally, although patterns in results were consistent across sites, they are limited to these areas and may not generalize to all settings. Applicability of findings to other areas, therefore, remains a topic for investigation.

## References

- Blake, J., & Davis, K. (1964). Norms, values, and sanctions. In R. Faris (Ed.), *Handbook of modern sociology*. Chicago, IL: Rand-McNally.
- Cancian, F. (1975). Introduction to the study of norms and strategies for describing and measuring norms. In F. Cancian (Ed.), *What are norms?* (pp. 1–26). New York, NY: Cambridge.
- Cohen, J. (1988). *Statistical power for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Hall, T., & Roggenbuck, J. (2002). Response format effects in questions about norms: Implications for the reliability and validity of the normative approach. *Leisure Sciences*, 24, 325–337.

- Heywood, J. (1996a). Conventions, emerging norms, and norms in outdoor recreation. *Leisure Sciences, 18*, 355–363.
- Heywood, J. (1996b). Social regularities in outdoor recreation. *Leisure Sciences, 18*, 23–38.
- Heywood, J. (2002). The cognitive and emotional components of behavior norms in outdoor recreation. *Leisure Sciences, 24*, 271–281.
- Heywood, J. (2011). Institutional norms and evaluative standards for parks and recreation resources research, planning, and management. *Leisure Sciences, 33*, 441–449.
- Manning, R. (2007). *Parks and carrying capacity: Commons without tragedy*. Washington, DC: Island Press.
- Manning, R. (2011). *Studies in outdoor recreation: Search and research for satisfaction*. Corvallis, OR: Oregon State University Press.
- Manning, R., & Freimund, W. (2004). Use of visual research methods to measure standards of quality for parks and outdoor recreation. *Journal of Leisure Research, 36*, 557–579.
- Manning, R., Valliere, W., Wang, B., & Jacobi, C. (1999). Crowding norms: Alternative measurement approaches. *Leisure Sciences, 21*, 97–115.
- Needham, M., Rollins, R., Ceurvorst, R., Wood, C., Grimm, K., & Dearden, P. (2011). Motivations and normative evaluations of summer visitors at an alpine ski area. *Journal of Travel Research, 50*, 669–684.
- Needham, M., Tynon, J., Ceurvorst, R., Collins, R., Connor, W., & Culnane, M. (2008). *Recreation carrying capacity and management at Pupukea MLCD on Oahu, Hawaii*. Corvallis, OR: Oregon State University.
- Opp, K. (1982). The evolutionary emergence of norms. *British Journal of Social Psychology, 21*, 139–149.
- Roggenbuck, J., Williams, D., Bange, S., & Dean, D. (1991). River float trip encounter norms: Questioning the use of the social norms concept. *Journal of Leisure Research, 23*, 133–153.
- Shelby, B., Vaske, J., & Donnelly, M. (1996). Norms, standards, and natural resources. *Leisure Sciences, 18*, 103–123.
- Vaske, J. (2008). *Survey research and analysis: Applications in parks, recreation and human dimensions*. State College, PA: Venture.
- Vaske, J., Shelby, B., Graefe, A., & Heberlein, T. (1986). Backcountry encounter norms: Theory, method, and empirical evidence. *Journal of Leisure Research, 18*, 137–153.
- Vaske, J., & Whittaker, D. (2004). Normative approaches to natural resources. In M. Manfredi, J. Vaske, B. Bruyere, D. Field, & P. Brown (Eds.), *Society and natural resources: A summary of knowledge* (pp. 283–294). Jefferson, MO: Modern Litho.