Journal of Leisure Research
2011, Vol. 43, No. 2, pp. 246-267

Generalizability of a Cluster Solution Over Time: Results from a Visitor Study at the Kennecott National Historic Landmark

Peter J. Fix
University Alaska Fairbanks

Stephen C. Taylor
Fairbanks North Star Borough
Parks and Recreation Department

Abstract

This study investigated the generalizability across time of a cluster analysis that identified five groups with differing motivations for visiting a National Historic Landmark. Replication of the initial study after a one-year time period suggested the same visitor types (i.e., clusters) were present in both study years; however, the relative size of the clusters varied by up to ten percent. A group defined by motivations scores in the middle of the 5-point scale exhibited weak recovery in the second year. Two scales constructed for this study exhibited high reliability in both study years, but low generalizability across time. Results provide implications for assessing trends, developing visitor typology-based management standards, and the development of scales.

KEYWORDS: Cluster analysis, cross-validation, generalizability, REP scales

Funding for this study was provided by the National Park Service, Wrangell – St. Elias NP and the North and West Alaska Cooperative Ecosystem Studies Unit. Megan Richotte, NPS, WRST, assisted in developing relevant management actions to be included on the survey. Michael Manfredo provided a valuable suggestion to use GT. Jerry Vaske provided an insightful review of a draft of the manuscript. We would like to thank anonymous reviewers for helpful suggestions, especially to use cross-validation. All errors are the responsibility of the authors.

Correspondence concerning this article should be addressed to Peter Fix, Department of Resources Management, School of Natural Resources and Agricultural Sciences, Box 757200 University of Alaska Fairbanks, Fairbanks, AK 99775. pjfix@alaska.edu, (907) 474-6926.

Introduction

Applied recreation research attempts to inform management decisions (Manning, 1999). Important considerations include sampling representativeness, reliability of measurement instruments, validity of the results, and generalizability of findings. This research used cross-validation (Breckenridge, 2000) and Generalizability Theory (GT) (Strube, 2000) to explore the generalizability of a study of visitor motivations at the Kennecott National Historic Landmark (KNHL) within Wrangell St. Elias National Park and Preserve, Alaska.

Applied research incorporates existing theories to address a particular problem (Kerlinger & Lee, 2000; Vaske, 2008). Many recreation studies advance theories and contribute knowledge to the field, but they are often embedded in a role to investigate a specific issue or management problem and provide recreation managers with the best possible information to guide or assist planning and policy decisions. For example, studies of norms as a tool to assist in carrying capacity decisions explored measurement issues such as question format, starting point bias, and information bias (Manning, Lawson, Newman, Laven, & Valliere, 2002); recreation research regarding attitudes offered insight into the effects of attitude strength and certainty (Bright, 1997); research on conflict has explored interpersonal versus social values conflict (Vaske, Donnelly, Wittmann, & Laidlaw, 1995), but these studies took place in the context of providing information to management.

When incorporating research into management decisions, a critical question is whether the results generalize to the visiting population relevant to the decision. Low generalizibility might be the result of sampling error in one season and/or variability from year to year that is not captured within a season's sample. For example, an article printed in a popular outdoor magazine might attract a type of visitor not present at the time the study was conducted. Road construction/closures, weather, marketing efforts of local/regional tourism boards, competition for visitors (e.g., the opening of a new destination in the area) can also influence visitation rates or the type of visitor present (Jackson, White, & Gronn, 2001). Research results intended to support planning will have the highest level of utility when it is representative of the entire visiting population to be covered by the plan.

Generalizability is important for several reasons. First, research conducted at one point in time might represent a year or season of visitors, but might not represent the visiting population. This is especially critical in trend studies, where the population that visited during the sample period might exhibit change, but the overall population of visitors might remain the same. In this case, visitor characteristics might actually be remaining stable over time. This would have implications for the extent management should focus on current visitors versus the entire visiting population. Second, existing planning frameworks (e.g., Visitor Experience and Resource Protection (VERP), Experience Based Management (EBM), the Recreation Opportunity Spectrum (ROS), etc.), attempt to preserve a given type of setting and experience (e.g., remote setting with opportunities for solitude). Management prescriptions for one course of action might be inappropriate for some visitors. In this case, it would be useful to know the barriers to implementation. Third, all recreation planning frameworks (e.g., VERP, EBM, ROS) utilize indicators and standards, with indicators being social or resource conditions managers and visitors care about and standards the acceptable level of the indicators (Vaske, Whittaker, Shelby, & Manfredo, 2002). Research conducted with on-site visitors as the study population can assist in setting the appropriate standards (Manning, 2007). However, if the visitor population has year to year variability, a study representative of a particular point in time (e.g., a year or a season) might not provide the appropriate level of the standard.

Generalizability in Outdoor Recreation Research

The generalizability across settings, locations, and time of constructs such as crowding, norms, satisfaction, and motivations, has been examined using meta or comparative analyses (Donnelly, Vaske, Whittaker, & Shelby, 2000; Kuentzel, Laven, Manning, & Valliere, 2008; Laven, Manning, & Krymkowksi, 2005; Manfredo, Driver, & Tarrant, 1996; Vaske & Donnelly, 2002; Vaske, Donnelly, Heberlein, & Shelby, 1982; Vaske & Shelby, 2008; Williams & Vaske, 2003). These studies have found the respective constructs studied to be applicable under a wide range of situations. Such studies contribute to our understanding of the construct, allow for theory refinement, and provide evidence of validity (Vaske & Manning, 2008).

In contrast to studies that examined the generalizability of constructs, this study examined the generalizability of results of a specific study across a time period to which they should generalize. Specifically, this study explored generalizability of study results across time by examining cluster analysis results. This had two steps, first examining the stability[1] of the clusters, i.e., do the same clusters emerge over time, then assessing the management implication of the results at different points in time. Cluster analysis allows for identification and classification of subsets of individuals who share similar responses (Jackel & Wollscheid, 2007; Needham, Vaske, Donnelly, Manfredo, 2007; Shinew, Floyd, & Parry, 2004; Shores & Scott, 2007; Walker, Jackson, & Deng, 2008; Vaske, Needham, & Cline Jr., 2007). Such analyses can indicate the proportion of visitors with certain characteristics, and inform managers who will be impacted by certain decisions (Vaske, 2008). The applied value of cluster analyses depends on the stability of the cluster solution and the generalizability across the visiting population of the cluster outcomes and study recommendations. In the absence of multiple samples across time for comparing cluster solutions, a split half design can be used to assess stability (Hair & Black, 2000). Shores and Scott (2007) used the split sampling method and found their cluster solution to be stable within their data. However, results from cluster analyses might not generalize across time (McLaughlin & Paradice, 1980). Given the prevalence of cluster analysis, the generalizability of cluster outcomes across time deserves further exploration. Cluster analyses would be enhanced by cross-validation using different measures of the original constructs and longitudinal replications over appropriate time periods (Breckenridge, 2000, p. 283). Research reported here is a longitudinal replication of our initial study. No site changes were made, and, thus, replication of results was expected. Generalizability involved cross-validation and generalizability theory. Cross-validation measures the degree of replication of results and differs from the split sample approach as it makes multiple comparisons and optimizes use of existing data (Mosteller & Tukey, 1968). Generalizabilty theory provides a framework to identify variation across time that might be the cause of inconsistent replication of results.

Cross-validation

Cross-validation is concerned with estimating the error of prediction of a model (Efron & Gong, 1983; Efron & Tibshirani, 1997). Although there are several methods of cross-validation (e.g., test set, leave one out, k folds) (Kohavi, 1995), this study assessed replication of cluster solutions, i.e., if new clusters are similar to existing clusters (Breckenridge, 1989; Breckenridge, 2000; Morey, Blashfield, & Skinner, 1983). In general, these studies partition data into two samples. Cluster analysis is conducted on both samples, and the data from one sample reclassified into the clusters of the other sample based on the classification rule used to create the clusters. The consistency of replication is then compared.[2] For example, if minimum squared Euclidean distance was used to create clusters for two samples, A and B, for each observation in A, the observation in sample B with the minimum squared Euclidean distance is identified (termed "nearest neighbor") and its cluster membership noted. This process can then be repeated where the nearest neighbor in A is identified, followed by the nearest neighbor in B (Figure 1). The adjusted Rand statistic (Hubert & Arabie, 1985) provides an index of replication (adjusting for chance agreement), ranging between zero and one, with zero indicating the two clusters are not similar and one indicating the clusters are identical. The adjusted Rand statistic can be calculated for several comparisons, for example recovery of actual cluster assignments, i.e., to what extent a known group's membership is recreated in a separate analysis, or replication of two separate clusterings of subsets of the data.

Generalizability Theory

Classical test theory models an observed score as consisting of a true score and measurement error and only allows the estimation of one type of error variance at a time (Shavelson, Webb, & Rowley, 1989). Thus, in classical test theory if there are multiple sources of error variance, they are all combined into one error term and the effects of different sources of variation cannot be isolated. Unlike classical test theory, variance in GT is composed of multiple sources of both systematic and random components with varying magnitudes and interactions. In contrast to classical test theory, GT allows us to understand how different sources of variation affect observed scores. GT assumes a universe score which has variability associated with each study component or "facet"[3] (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). The researcher-specified facets of the study define measurement conditions that might affect obtained score variance. Systematic variability is associated with facets included in the GT model (e.g., time, setting, individuals taking the test, survey administrator). By analyzing systematic variance in relation to the various facets of the study, GT can determine the degree of accuracy by which results generalize across possible occurrences, settings, and populations (Shavelson et al., 1989). Recreation applications of GT have assessed the generalizability of satisfaction measures across time, setting, and context (Schomaker & Knopf, 1982), a single item from the Recreation Experience Preferences (REP) scales across setting (Manfredo, 1984), and REP scales across time, format, and subjects (Williams, Ellis, Nickerson, & Shafer, 1988).

This research assessed the generalizability of a study at KNHL that used cluster analysis to link motivations for visiting (measured by the REP scales) with visitor characteristics, such as activity participation, length of stay, and demographics, and management preferences. The initial study results, based on cluster solutions, were derived from data collected by an on-site survey during the summer of 2004. A replication of the survey during the summer of 2005 provided data for assessing the generalizability of the cluster solution. Given that no deliberate management changes occurred in KNHL between 2004 and 2005, and economic conditions (e.g., gas prices, employment rates, etc.) were stable, visitors should have similar motivations, characteristics, and activity patterns, and, thus, similar clusters will re-emerge in 2005. We hypothesize:

H1 Clusters that emerge in 2004 will also be present in 2005.
H2 Variance in the GT analysis will be explained by cluster rather than time.

Methods

Visitors to the KNHL were randomly sampled over two consecutive summers. Days were divided into two time blocks and the following two stage cluster sample formula (Cochran, 1977) used to determine the number of time blocks to sample and the number of visitors to sample in each time block in 2004.

$$m = \frac{s_2}{\sqrt{s_1^2 - \frac{s_2^2}{M_i}}} \qquad n = \frac{s_1^2 - \frac{s_2^2}{M_i} + \frac{s_2^2}{m}}{V(\hat{\mu})}$$

where: n = the number of clusters (i.e. time blocks) sampled; m = the number of elements (i.e. visitors) sampled per cluster, s_1^2 = the estimated variance between clusters, s_2^2 = the estimated variance within a cluster, M_i = the average number of elements per cluster, $V(\hat{\mu})$ = the variance of the estimated mean of the population. As no variance estimates from an Alaska National Park were available, we used an estimate of variance from a Rocky Mountain National Park study (Stewart, Fix, & Manfredo, 2004). Assuming a 70% response rate, we estimated six people should be sampled in each of 53 time blocks and sampled every third visitor exiting the KNHL. Funding constraints necessitated a consolidated approach in 2005. Three one-week time periods were randomly selected to be sampled between noon and 6 p.m. The days not included in these time blocks were sampled at two times each day, three hours apart, with the starting time varying each day. Each sampled visitor completed an on-site survey related to their motivations, trip characteristics, and preferences for management actions. The first sample (t_1) was taken from June 11 through September 6 (Labor Day), 2004. The second sample (t_2) was taken from July 8 through September 5 (Labor Day), 2005. Respondents in t_1 did not differ on key variables by the time of summer they visited, so the later start of t_2 did not present a great concern.[4] Visitors 18 years of age and older completed the on-site questionnaire as they departed the Kennecott mill town.

Visitor motivations were recorded on 18 statements, representing eight motivational domains hypothesized to be relevant to Kennecott. Scale items from seven of the eight domains were selected from Driver's (1983) master list of REP scales (Table 1). Scale items for the eighth domain, history, were developed by the authors, in consultation with park staff, to address a hypothesized motivation of history due to the historic nature of the Kennecott mill town. Multiple scale items from each domain were included to allow a verification of their reliability during the data analysis stage. To measure motivations for taking the trip to KNHL, responses were recorded on a five-point scale ranging from "not at all important" to "extremely important" (Manfredo, et al., 1996).

Respondents were also provided a list of 14 activities (plus an option for an open-ended response) and asked to indicate which activities they participated in and which was their primary activity. Several characteristics regarding their trip were measured, including: length of stay, prior visitation, mode of travel to Kennecott, travel companions, types of information used, and subject about Kennecott that was of most interest. Demographic information was also collected.

A nonresponse test was conducted. In t_1 nonrespondents were asked two questions: if they had visited the Kennicott Valley[5] before and preference for further structure stabilization in the Kennecott mill town. We included these questions as they might represent meaningful differences among respondents and they could quickly be asked, even if respondents were in a hurry. Four characteristics of both the respondents and nonrespondents were noted: group size, number of children in group, time of day (only as morning vs. afternoon time block in t_1) and gender. In t_2, the same variables were included in the nonresponse test except prior visitation as the majority of the visitors in t_1 were first time visitors.

Cluster Identification

Cluster solutions were derived using the same protocol for both samples. The reliability of reported motivation scores was analyzed by REP scale domains using Pearson's *r* correlation coefficient for domains with two scale items and Cronbach's alpha for domains with three or more scale items. Sufficient reliability was assumed to occur with coefficients of 0.7 or higher (Garson, 2002). Cluster analysis was conducted on the means of the REP scales. Following recommendations by Hair and Black (2000), potential outliers were identified using hierarchical cluster analysis and the profiles of potential outliers examined. Box plots of the clustering variables' means were used to confirm whether these cases were outliers. After identification and removal of outliers, final clusters were identified using K-means cluster analysis. Cluster analyses were run with 3, 4, 5, and 6 groups. The final determination of the appropriate group number was based on the variance between clusters versus within clusters, as indicated by the F values; the distinctness of the clusters, determined by comparison of the mean scores; and the relative number of cases within each group. Themes for the clusters in each year were then developed using a two step process. First, the dominant REP scale scores were identified. Second, for each year, the clusters' relationship with other variables was examined

Table 1

Motivation Scale Items, Means and Reliabilities, 2004 and 2005 Kennecott National Historic Landmark Visitor Survey

Motivation domain	2004			2005		
Scale items	**M**	**SD**	**Reliability**	**M**	**SD**	**Reliability**
Escape Physical Pressure			.86 [a]			.83 [a]
To experience tranquility	3.73	1.13		3.65	1.24	
To be away from crowds of people	3.80	1.14		3.73	1.20	
To experience natural quiet[c]	3.93	1.03		4.04	1.04	
History			.73 [b]			.70 [b]
To learn about the history of the area	3.95	1.08		4.01	1.03	
To be in a historical setting	3.81	1.12		3.79	1.17	
Exercise/Physical Fitness			.90 [b]			.85 [b]
To get exercise	3.61	1.25		3.56	1.30	
To feel good after being physically active	3.76	1.15		3.70	1.27	
Enjoy Nature			.83 [a]			.84 [a]
To enjoy the sounds and smells of nature	4.07	0.97		4.08	.94	
To be in a natural setting	4.20	0.93		4.21	.92	
To observe wildlife	3.81	1.12		3.91	1.09	
Learning			.71 [b]			.72 [b]
To learn more about nature	3.43	1.06		3.60	1.11	
To learn more about the ecology of the area	3.53	1.10		3.60	1.11	
Family			.69 [b]			.68 [b]
To bring your family close together	2.38	1.39		2.29	1.42	
To be with family	2.91	1.57		2.95	1.63	
Companionship			.56 [b]			.57 [b]
To be with friends	2.74	1.48		2.59	1.62	
To be with others who enjoy the same things you do	3.13	1.35		3.17	1.49	
Family/Companionship[d]			.74 [a]			.72 [a]
Creativity			.52 [b]			.52 [b]
To gain a new perspective on life	2.64	1.30		2.59	1.39	
To do something creative such as paint, sketch, or photograph	2.55	1.47		2.54	1.57	

[a] Cronbach's Alpha
[b] *r*
[c] Included at the request of the National Park Service
[d] Includes the family and companionship scale items combined

(i.e., profiling) (Hair & Black, 2000). Profiling is commonly used to further identify characteristics of clusters and establish face validity of cluster solutions (Cha, McCleary, & Uysal, 1995; Coupal, Bastian, May, & Taylor, 2001; Harris, 2004; Hautaluoma & Brown, 1978; Hudson & Ritchie, 2002; Loker-Murphy, 1996; Manfredo & Larson, 1993; May, Bastain, Taylor, & Wipple, 2001; Needham et al., 2007; Oh, Ditton, Anderson, Scott, & Stoll, 2005; Swanson, Vande Kamp, & Johnson, 2002). Finally, the cluster outcomes from t_1 and t_2 were compared to determine if similar clusters themes emerged across years. This involved two steps. First, the relative importance of the mean domain scores within and among the clusters was compared. This was done by the authors of the study and students in graduate recreation management classes. Second, we compared profiles of clusters across the two years. The dominant REP scale means and cluster profile were used to develop a theme for each cluster. Clusters with similar dominant means and cluster profiles were assigned the same theme and cluster number.

Cross-validation

Following Breckenridge (2000), we examined the nearest neighbor of each case in t_1 with the data in t_2. Squared Euclidean distance was the algorithm utilized in the cluster analysis, and was the method used to determine the nearest neighbor. Observation x_i's nearest neighbor in t_2 is the case with the minimum squared Euclidean distance from x_i. A test for recovery was conducted by examining the nearest neighbor in t_2 of all observations in t_1. This test assumes that the clusters in t_1 represent the true clusters and the data in t_2 should replicate those clusters. The Hubert and Arabie adjusted Rand statistic (Hubert & Arabie, 1985; Milligan & Cooper, 1986) was used as a measure of the extent of recovery of the data's structure. The adjusted Rand statistic compares all possible pairs of cases within a sample and their nearest neighbors across clustering sequences, in our case the samples in t_1 and t_2, to determine if they are in the same clusters, different clusters, or some mix across the two cluster analyses. For example, if there are ten observations in the cluster analysis, there would be 45 possible pairs, e.g., x_1 and x_2, x_1 and x_3, x_1 and x_4, etc. Specifically, the following formula from Park and Jun (2009) was used:

$$RI_{adj} = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)}$$

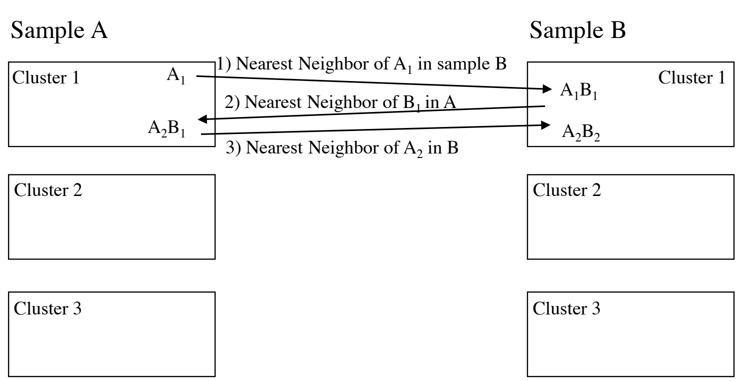
Where a is the number of pairs of objects in the same cluster in t_1 and the same cluster in t_2, b is the number of pairs in the same cluster in t_1 but not in the same cluster in t_2, c is the number of pairs not in the same cluster in t_1 but in the same cluster in t_2, and d is the number of pairs in different clusters in t_1 and different clusters in t_2.

However, the Rand statistic (and, thus, the adjusted Rand statistic) treats two pairings of cluster assignments that are in different clusters across both samples as a "correct assignment" regardless of whether there is any consistency in the pairings (e.g., if for observation 1 the clusters for t_1 are 1 and 3 and the clusters for t_2 are 2 and 4, and for observation 2 the clusters for t_1 are 1 and 3 and for t_2 are 2 and 5, the pairs would be counted as "different clusters in both samples" even though the clusters are varying). As our research is concerned with specific cluster assignment and because the adjusted Rand statistic does not describe the clusters of mismatched pairs, simple frequencies were conducted on the matched and mismatched clusters.

A test for replication was also conducted. This was conducted following the sequence of Figure 1, e.g., a "double cross-validation." Breckenridge (2000) defines two points as being reciprocal nearest neighbors (RNN) when A_1 and A_2B_1 are in the same cluster and A_1B_1 and A_2B_2 are in the same cluster. However, in this study the cluster assignment is of importance as clusters that were judged to have a similar theme were assigned the same cluster number (e.g., cluster 1 in t_1 should match cluster 1 in t_2) and, therefore, our results were examined with an added constraint that nearest neighbors be in the same cluster across samples (RNNsc). The percent of RNNsc was calculated as well as the frequencies of the clusters that were RNNsc and cluster combinations that were not.

Figure 1

General Steps for Cross-validation used to Compare Cluster Membership in 2004 and 2005 Data from a Visitor Survey at the Kennecott National Historic Landmark



Generalizability Assessment

GT required us to first assess if similar clusters emerged in t_1 and t_2 and then assume respondents in these clusters are homogenous, i.e., a respondent classified into Cluster 1 in t_1 is equivalent to a respondent classified into Cluster 1 in t_2 (cluster numbers were coded to match based on their profile). Thus, we used a GT with a nested design (Salter, 2003; Stuhlman, Daniel, Dellinger, Denny, & Powers, 1999). This approach was appropriate because the goal was to test the generalizability of the cluster solutions. The dependent variable in the GT (i.e., the REP scale scores) is a composite of two or three questions, however GT is applicable to composite scores (Cronbach et al., 1972). Similar to Strube (2000), we used one-facet, where the variation in motivation scores (Y_m), was composed of the variation associated with cluster (*c*), time (*t*), and the variation in the interaction of cluster by time and random error (*ct,e*).

$$\sigma^2(Y_m) = \sigma^2(c) + \sigma^2(t) + \sigma^2(ct,e)$$

The equation was estimated with ANOVA type III sum of squares due to the unequal ns among the clusters. The variance components were estimated with the following equations (Strube, 2000):

$\hat{\sigma}^2(c) = [MS(c) - MS(ct)]/n(t)$
$\hat{\sigma}^2(t) = [MS(t) - MS(ct)]/n(c)$
$\hat{\sigma}^2(ct) = MS(ct)$

This method can result in negative estimated variance components (Shavelson et al., 1989). The one-facet nature of this study does not require any of the potentially negative variance components to be used in computations, thus any negative variance component values were set to zero for reporting purposes (Cronbach et al., 1972). Generalizability tests were conducted for each motivation item, providing insight into the generalizability of each motivation in the cluster outcome.

Results

Three hundred fifty-one visitors were contacted and 233 surveys were completed in t_1 (66% response rate); 300 visitors were contacted in t_2, resulting in 225 completed surveys (75% response rate). The nonresponse test did not reveal significant differences (at $p = .05$) in any of the 11 nonresponse tests.[6] Seventeen of the 18 REP variables in t_1 and 15 of the 18 in t_2 had skewness indices within an acceptable range of -1 and +1 (Morgan, Leech, Glockner, & Barret, 2004). A Levene's test for equality of variance prior to reliability analysis revealed five of the eight scales in t_1 and four of the eight in t_2 had scale items with homogenous variance ($p = .01$). In 2004, the scales with three variables had mixed results among the comparisons, and in 2005, one of the scales with three variables had mixed results. As the tests used in this paper are extremely robust under violations of normality and homogenous variance assumptions, especially with large sample sizes with relatively equal ns (Harris, 2001, pg. 450; Morgan et al., 2004; Vaske, 2008), these violations should not impact the interpretation of our test statistics. In t_1, five of the eight domains met the criteria of reliability > = 0.70: exercise/physical fitness, learning, enjoy nature, escape physical pressure, and history. The family domain had reliability of .69 and companionship .56. Given the observed social nature of Kennecott visitors, we hypothesized there to be a social motive represented by a combination of the family and companionship scale items, consistent with other research (Coupal et al., 2001; Legare & Haider, 2008). The combined family and companionship scale demonstrated adequate reliability. For consistency, the same domains were used in t_2, with reliability being similar to t_1 (Table 1). Creativity had low reliability and was excluded from the cluster analysis. The hierarchical cluster analysis revealed seven cases in t_1 and eight cases in t_2 that could be potential outliers; a box plot confirmed their domain scores (which were lower than the means) were outliers. These cases were excluded from the k-means cluster analysis. Item non-response further decreased the cluster sample sizes to 206 (t_1) and 198 (t_2). As only 15% of the visitors in 2005 had previously visited the Kennicott Valley, we assumed 2004 and 2005 were independent samples and, therefore, used two-factor ANOVA without adjusting for repeated measures.

Cluster Solution

For t_1 and t_2, the five-cluster solution resulted in the most distinct groups, with adequate combined F values (F = 286 and 332 in t_1 and t_2, respectively), that were not dominated by one variable. The relative importance of each cluster's defining motivations was consistent in both years (Figure 2). The cluster outcomes were further validated by differences in the defining activities (Table 2) and primary subjects of interest (Table 3). Cluster 1 was a highly active group that enjoyed solitude, nature, and exercise while placing little importance on history ("Outdoor Enthusiasts"). Cluster 2 visitors placed importance on all aspects of the park, especially interpretive programs, wildlife viewing, and hiking ("Park Experience"). Cluster 3 placed a high importance on the history of the Kennecott compared to other motivations ("History Buff"). Cluster 4 visitors did not have any dominant motivation ("Generalists"). Finally, Cluster 5 placed a high importance on history, nature, and learning and a low importance on family/companionship ("Tourist").

The cluster solutions differed in the distribution of respondents among the clusters across years (χ^2[4, n = 404] = 13.75, p = .008). Clusters that exhibited a lower representation in t_2 consisted of Cluster 3 (22.3% to 12.1%) and Cluster 5 (19.9% to 14.6%). Clusters 1 and 4 each increased roughly 7% to 22.7% and 17.2% respectively in t_2. Finally, Cluster 2 remained steady, representing just above 30% of each sample.

Cross-validation

We used the sample from t_1 as the baseline for recovery and replication (n = 206). Relative to recovery, hypothesis 1 was partially supported as 125 (61%) of t_1's nearest neighbors in t_2 were in the same cluster. The adjusted Rand statistic was .34, lower than Breckenridge's (2000) result of an adjusted Rand of .7 for 5 clusters of 6 variables with random noise added to the data. However, it is consistent with Breckenridge's adjusted Rand statistic of .3 when random variables were added to the data, perhaps a more realistic comparison for empirical data. With respect to which clusters were matched, Cluster 2 had 77% of its NNs in t_2 match, and Cluster 1 and 3 had 64% and 61% of their NNs match, respectively. Cluster 5 had approximately 50% of its NNs match, while cluster 4 had only 29% of its NNs match (Table 4). For Clusters 1, 2, 3, and 5, there was no other cluster combination that would have resulted in greater recovery (Table 5). Recovery of Cluster 4, however, could have improved to 38% if it were matched to Cluster 1 (although, given its low n this only represents two additional cases). Cluster 1 in t_2 was often the nearest neighbor for the other clusters (i.e., Clusters 2 through 5 in t_1).

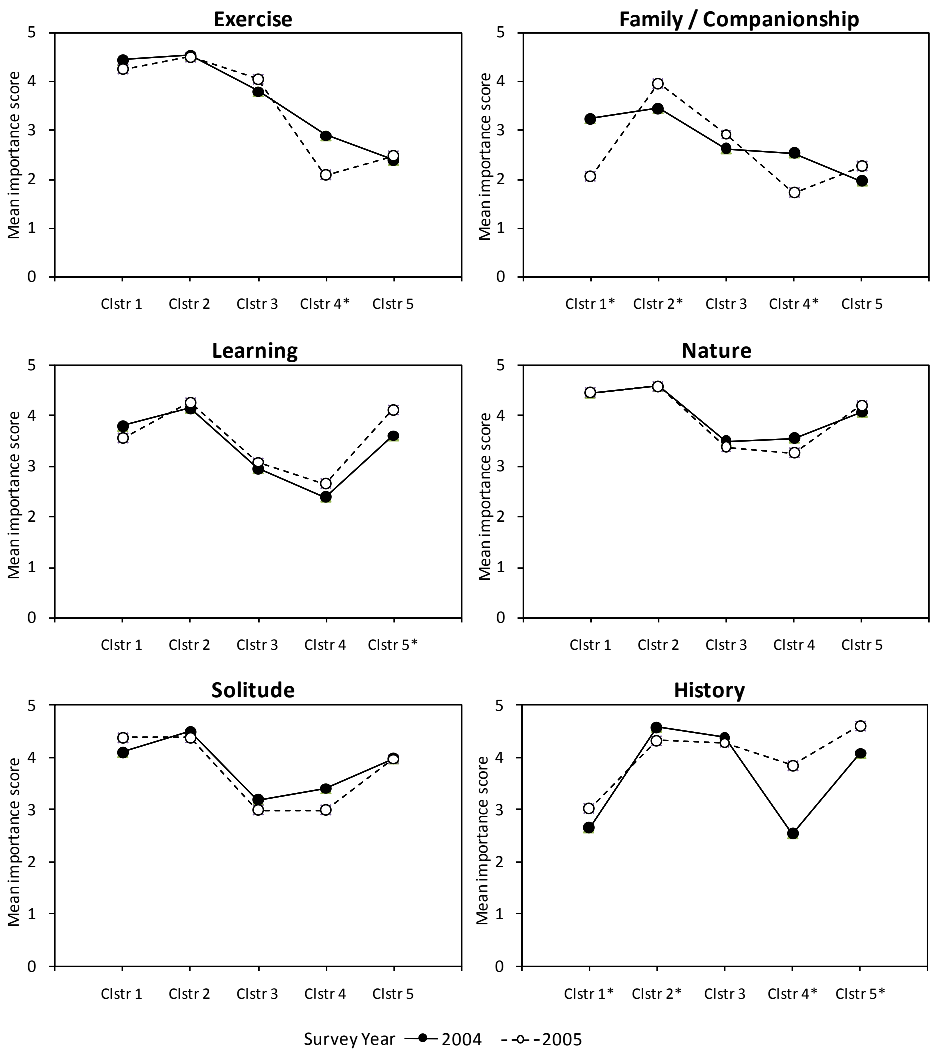


Figure 2. Motivation Score Profile Plots by Cluster Membership, 2004 and 2005 Kennecott National Historic Landmark Visitor Survey. Estimated importance scores based on 5-point response scale where 1 = "not at all important" and 5 = "extremely important." Clstr = Cluster. ns for Clusters 1, 2, 3, 4, 5 in 2004 were 33, 65, 46, 21, and 41, respectively and in 2005, 45, 66, 24, 34, and 29, respectively (χ^2 [4, n = 404] = 13.75, p = .008). * significant difference between 2004 and 2005 at $p < .05$ (measured with t test).

Table 2

Primary Activity Categories of Respondents by Motivation-Based Cluster Membership, 2004 and 2005 Kennecott National Historic Landmark Visitor Survey

		Primary activity category				
Cluster	**n**	**Active outdoors[a]**	**Nature enjoyment[b]**	**History & interpretation[c]**	**Sightseeing[d]**	**Other**
			2004 survey[e]			
Cluster 1	28	82.1%	7.1%	3.6%	0.0%	7.1%
Cluster 2	46	37.0%	13.0%	30.4%	17.4%	2.2%
Cluster 3	38	39.5%	13.2%	28.9%	18.4%	0.0%
Cluster 4	20	50.0%	5.0%	15.0%	10.0%	20.0%
Cluster 5	36	30.6%	5.6%	44.4%	11.1%	8.3%
			2005 survey[f]			
Cluster 1	42	61.9%	11.9%	11.9%	7.1%	7.1%
Cluster 2	64	40.6%	7.8%	21.9%	23.4%	6.3%
Cluster 3	21	33.3%	9.5%	23.8%	33.3%	0.0%
Cluster 4	32	18.8%	6.3%	53.1%	18.8%	3.1%
Cluster 5	28	21.4%	10.7%	42.9%	17.9%	7.1%

Note. Cell entries are the percent of respondents in each cluster selecting an activity within this category as their primary activity.
[a] Includes mountaineering, climbing, hiking, backpacking, rafting, and biking
[b] Includes wildlife viewing, nature walks, camping, and fishing
[c] Includes interpretive programs and exploring the historic mill town
[d] Sightseeing and flight seeing
[e] Activity categories were significantly different among clusters (χ^2[16, n = 168]= 40.7, p = .001).
[f] Activity categories were significantly different among clusters (χ^2[16, n = 187] = 34.6, p = .005).

The test for replication also partially supports hypothesis 1. A reciprocal match occurs when the nearest neighbors within a respective time are in the same cluster, 151 (73%) of the observations are reciprocal nearest neighbors. Adding the criteria that all nearest neighbor assignments must be in the same cluster (RNNsc), 111 (54%) of the observations are nearest neighbors. Examining just the matched cases with the more stringent criteria (i.e., the second analysis), observations belonging to Cluster 2, Park Experience, had the greatest number of reciprocal nearest neighbors (n=49), both as a percent of the 111 RNNsc and as a percent of the number of observations in the cluster (75%). Observations from Cluster 4, Generalists, had the fewest RNNsc (5), both as a percent of the 111 RNNsc and as a percent of the number of observations in the cluster (24%). With respect to the clusters of the mismatched RNNsc, mismatched pairs with no more than two clusters included in the four-cluster sequence (e.g., the pattern of cluster memberships for t_1 x_1 (A), t_2 NN of A (B), t_1 NN of B (C), and t_2 NN of C might be 1,1,1,2, respectively) were examined. Clusters with combinations of Cluster 1 (Outdoor Enthusiasts) in t_1 and Cluster 2 (Park Experience) in t_2 comprised 12 of the 94 mismatched cases, this was followed by 10 cases with Cluster 5 (Tourist) in t_1 and Cluster 4 (Generalist) in t_2 (Table 6). The clusters of 1 and 2 seem to fit together often and Cluster 1 showed up often as the cluster of the NN for Clusters 2, 3, and 4 in t_1.

Table 3

Primary Subject of Interest[a] of Respondents by Motivation-Based Cluster Membership, 2004 and 2005 Kennecott National Historic Landmark Visitor Survey

		Primary subject of interest				
Cluster	**n**	**Wildlife**	**History**	**Ecology**	**Geology**	**Other**
			2004 survey[b]			
Cluster 1	27	11.1%	14.8%	37.0%	25.9%	11.1%
Cluster 2	45	8.9%	64.4%	4.4%	15.6%	6.7%
Cluster 3	34	2.9%	88.2%	0.0%	8.8%	0.0%
Cluster 4	20	5.0%	60.0%	5.0%	25.0%	5.0%
Cluster 5	34	5.9%	70.6%	5.9%	17.6%	0.0%
			2005 survey[c]			
Cluster 1	42	16.7%	40.5%	9.5%	26.2%	7.1%
Cluster 2	62	6.5%	61.3%	4.8%	16.1%	11.3%
Cluster 3	23	0.0%	91.3%	0.0%	8.7%	0.0%
Cluster 4	33	3.0%	60.6%	6.1%	27.3%	3.0%
Cluster 5	27	3.7%	81.5%	3.7%	7.4%	3.7%

Note. Cell entries are the percent of respondents in each cluster selecting each subject as their primary subject of interest.
[a] Respondents were provided a list of these subjects and asked to identify the subject that interested them the most.
[b] Subjects of interest were significantly different among clusters (χ^2[16, n = 160] = 52.7, p < .001).
[c] Subjects of interest were significantly different among clusters (χ^2[16, n = 187] = 29.8, p = .019).

Table 4

Recovered Clusters and Reciprocal Nearest Neighbors (Same Cluster) of t_1, 2004 and 2005 Kennecott National Historic Landmark Visitor Survey

Cluster in t_1	**n**	**Number of Recovered Clusters (%)**	**Number of RNNsc (%)**
1	33	21 (64%)	15 (45%)
2	65	50 (77%)	49 (75%)
3	46	28 (61%)	24 (52%)
4	21	6 (29%)	5 (24%)
5	41	20 (49%)	18 (44%)

Notes. t_1 is 2004. RNNsc is reciprocal nearest neighbor in the same cluster. The percent is based on the clusters' n.

Table 5

Cluster Pairs of Non-recovered Clusters of t_1, 2004 and 2005 Kennecott National Historic Landmark Visitor Survey

Cluster in t_1	Cluster in t_2	Number of cases
1	2	11
1	5	1
2	1	6
2	3	7
2	5	2
3	1	9
3	2	3
3	4	4
3	5	2
4	1	8
4	2	1
4	3	6
5	1	9
5	4	12

Note. t_1 is 2004 and t_2 is 2005.

Table 6

Most Common Nearest Neighbors of Mismatched Reciprocal Nearest Neighbors, 2004 and 2005 Kennecott National Historic Landmark Visitor Survey

Cluster in t_1	Cluster in t_2	Number of cases
1	2	12
2	1	5
2	3	7
3	1	5
4	1	8
5	4	10

Note. t_1 is 2004 and t_2 is 2005.

Generalizability of Motivation

Consistent with hypothesis 2, time accounted for a negligible amount of variance when the six defining motivations were tested in the GT framework (Table 7). The cluster components of variation were high for nature, exercise, solitude, and learning (94%, 93%, 90%, and 89%, respectively). Alternatively, the family/companionship and history domains exhibited large components of residual variation (46.4% and 26.5% respectively). The relatively large *ct,e* component for the history and family/companionship domains is also evident in the change in these two variables' ratio of variance among clusters to the variance within clusters (i.e., *F* value calculated with ANOVA). The results show in t_1, history had high variance among clusters ($F[4, 201] = 100.22$, $p = <.001$) and low variance among clusters in t_2 ($F(4, 191) = 27.70$, $p = <.001$), whereas family/companionship increased from t_1 to t_2 ($F[4, 201], = 18.82$, $p = <.001$ and $F[4, 191] = 78.94$, $p = <.001$, respectively).

Table 7
Variance Components for Motivations, 2004 and 2005 Kennecott National Historic Landmark Visitor Survey

	Estimated variance component[a, b]			**Percent of variation**		
Motivation domain	Cluster (*c*)	Year (*t*)	*ct, e*	*c*	*t*	*ct, e*
Exercise	36.97	0.00	2.88	92.8%	0.0%	7.2%
Family/Companion	14.38	0.00	12.45	53.6%	0.0%	46.4%
Learning	16.09	0.10	1.88	89.0%	0.6%	10.4%
Nature	11.30	0.00	0.67	94.4%	0.0%	5.6%
Solitude	13.49	0.00	1.55	89.7%	0.0%	10.3%
History	18.73	1.03	7.13	69.6%	3.8%	26.5%

Note. ct, e is the interaction of cluster by time and random error.
[a] Sum of squares should be interpreted based on a 5-point response scale.
[b] Negative variance component values set to 0.

This contrasts other variables that had *F* values that did not differ by more than 26.2 between t_1 and t_2. Examining a plot of cluster by time for each of the motivations reveals most of the motivation domains have lines that are near parallel and close together verifying the high component of variation due to cluster membership (Figure 2). The plots also reveal that Cluster 4 tended to exhibit greater variation in motivation scores across time than the other four clusters, especially for family/companionship and history.

Conclusions and Discussion

Temporal Stability of the Cluster Outcome

Four of the five clusters (1, 2, 3, and, to a slightly lesser extent, Cluster 5) were present in both years. This supports the cluster groups as being representative of the true structure of the visitors. However, Cluster 4 (Generalist) only had 24% of the clusters in t_2 as RNNsc. Cluster 4 had more RNNsc with cluster 1 in t_2 (8 vs. 5) than Cluster 4. With respect to Cluster 4, a visual examination of the domain means and its profile suggested its presence in both years. However, it appears this group shifted with respect to which cluster in t_1 its scores most resembled. Several explanations are possible. A small shift in REP scores of one or several of the members of Clusters 1 and 4 might have restructured the minimum squared Euclidean distances of Cluster 4 with respect to the other groups. In this scenario, Cluster 4 is a true grouping of visitors and the poor recovery is due to variation within members of the groups. As the cluster analysis was conducted on the data in their original value rather than using standardized scores (e.g., an ipsitive transformation), another possible explanation for the overlapping clusters is that better separation of clusters could be attained with an ipsitive transformation. However, we checked the data prior to analysis and determined respondents had relatively high variation in responses and using unstandardized variables was the best strategy. A post hoc comparison of the clustering variables' standard deviations across the clusters shows that Cluster 4's standard deviation was within the range of the other clusters (ranging from 0.68 to 1.12, with Cluster 4 = 0.84 in t_1 and ranging from 0.57 to 1.14, with Cluster 4 = 1.06 in t_2). Thus, it does not appear something was different about Cluster 4 with respect to scale variability. Nonetheless, the nature of this group should be examined through a cross-validation study comparing recovery between standardized and unstandardized scores.

The number of respondents in each of the clusters differed significantly across the two years. The change in the number of respondents in each cluster has implications for management. A program developed for a particular group might experience a 10% percent variation in attendance. Such fluctuations should be accounted for when developing standards to evaluate the success of a program. Potential year to year variation in cluster membership should be considered when monitoring trends over time. For example, Legare and Haider (2008) examined trends in visitation. Their comparison of cluster analyses from three points in time (1993, 1998, 2004) showed significant differences in the number in each of three clusters across time, including a directional change of 15% in one cluster. Our data showed a 10% shift across a one year time period, in which no management changes were made and the economic variables were stable. Recreation research needs to analyze multiple datasets and assess trends. However, we need to understand annual variation in visitation before trends can be assessed. Given the potentially intrusive nature of our research (i.e., a visitor survey) we might be at a disadvantage in this area (e.g., as opposed to monitoring the polar ice cap by satellite imagery). However, when possible, we should seek to design studies around multiple seasons to better capture year to year variation.

GT complemented cross-validation by identifying inputs to the cluster analysis that might contribute to recovery and replication. The large experience type component of variation for four of the six motivations used to cluster the respondents exhibited a high degree of generalizability across years. Furthermore, the time component of variation was minimal for all six motivations, signifying a relatively high degree of temporal stability of the cluster solutions. Of particular interest were the ct,e variance components in the family/companionship and history domains. While the cluster by time interaction cannot be separated from the random error in the ct,e component, Strube (2000) argues that a large ct,e component tends to signal an inadequate understanding of the facets that affect score variability.

The family/companionship and history domains represented a departure from the REP scale structure. Consistent with findings from other studies, questions from the family and companionship domains were combined for this study to represent a social motivation (Coupal et al., 2001; Legare & Haider, 2008). The history domain was created specifically for this study to reflect the unique characteristics of the study site. The REP scales were extensively tested during their development. Our results suggest caution should be used when deviating from the REP structure as certain scale items might vary together for a given study setting and time period, but their generalizibility across time might be low. Furthermore, the large residual component of variation evident in a generalizability study might not be obvious in the internal reliabilities calculated for a single sample. Reliability coefficients calculated in this study for the history and family/companionship domains were each above .70 across both samples, yet the generalizability of the cluster analysis results across the two time periods of this study was low. As other researchers have found combinations of REP questions that go across pre-set domains provide the best fit for their study settings (Coupal et al., 2001; Legare & Haider, 2008; Nyaupane, White, Budruk, 2006; Petrick, Backman, Bixler, & Norman, 2001; White, Virden, & Cahill, 2005), the generalizability of study findings across time should be further explored.

Applied research study results must possess some degree of generalizability over time if they are to be of utility to recreation managers and planners. In this study, four of the five clusters and their associated experience and activity characteristics exhibited strong recovery as measured by cross-validation, although their numbers fluctuated between years. This suggests the study captured a large component of the true structure of the visitors. However, when designing policies and actions, managers should allow for some oscillation in makeup of Cluster 4 over time and variation in the numbers of visitors in the other clusters. With these considerations, the results should provide a sound basis for management actions.

References

Breckenridge, J. N. (2000). Validating cluster analysis: consistent replication and symmetry. *Multivariate Behavioral Research, 35*, 261-285.

Breckenridge, J. N. (1989). Replicating cluster analysis: method, consistency, and validity. *Multivariate Behavioral Research, 24*, 147-161.

Bright, A. D. (1997). Attitude-strength and support of recreation management strategies. *Journal of Leisure Research, 29*, 363-379.

Cha, S., McCleary, K. W., & Uysal, M. (1995). Travel motivations of Japanese overseas travelers: A factor-cluster segmentation approach. *Journal of Travel Research, 34*, 33-39.

Cochran, W. G. 1977. *Sampling techniques*. New York: John Wiley & Sons.

Coupal, R. H., Bastian, C., May, J., & Taylor, D. T. (2001). The economic benefit of snowmobiling to Wyoming residents: A travel cost approach with market segmentation. *Journal of Leisure Research, 33*, 492-510.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: John Wiley & Sons, Inc.

Donnelly, M. P., Vaske, J. J., Whittaker, D., & Shelby, B. (2000). Toward an understanding of norm prevalence: A comparative analysis of 20 years of research. *Environmental Management, 25*, 403-414.

Driver, B. L. (1983). Master list of items for Recreation Experience Preference scales and domains. Unpublished Document. USDA Forest Service, Fort Collins, CO: Rocky Mountain Forest and Range Experiment Station.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician, 37*, 36-48.

Efron, B. & Tibshirani, R. (1997). Improvements on Cross-validation: the .632+ bootstrap method. *Journal of the American Statistical Association, 92*, 548-560.

Garson, D. G. (2002). *Guide to writing empirical papers, theses, and dissertations*. New York: Marcel Dekker, Inc.

Hair, W. C., & Black, J. F. (2000). Cluster analysis. In Grimm, L. G., and Yarnold, P. R. (Eds.) *Reading and understanding more multivariate statistics*. Washington, DC: American Psychological Association.

Harris, R. J. (2001). *A primer of multivariate statistics* (3rd ed.). Mahwah, NJ: Lawrence Earlbaum Associates.

Harris, S. H. (2004). A market survey of ecotourists in the Valdivian temperate forest ecoregion of Chile. Unpublished master's thesis, University of Alaska Fairbanks, Fairbanks, Alaska.

Hautaluoma, J., & Brown, P. J. (1978). Attributes of the deer hunting experience: A cluster-analytic study. *Journal of Leisure Research, 10*, 271-287.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*, 193-218.

Hudson, S., & Ritchie, B. (2002). Understanding the domestic market using cluster analysis: A case study of the marketing efforts of Travel Alberta. *Journal of Vacation Marketing, 8*, 263-276.

Jackel, M., & Wollscheid, S. (2007). Time is money and money needs time? A secondary analysis of time-budget data in Germany. *Journal of Leisure Research, 39*, 86-108.

Jackson, M., White, G., & Gronn White, M. (2001). Developing a tourist personality typology. Capitalising on Research. *Proceedings of the Eleventh Australian Tourism and Hospitality Research Conference* (pg.177-184). Canberra, ACT.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. Fort Worth, TX: Harcourt College Publishers.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, Quebec, Canada, pg. 1137-1143.

Kuentzel, W. F., Laven, D., Manning, R. E., & Valliere, W. A. (2008). When do normative standards matter most? Understanding the role of norm strength at multiple national park settings. *Leisure Sciences, 30*, 127-142.

Laven, D., Manning, R., & Krymkowski, D. (2005). The relationship between visitor-based standards of quality and existing conditions in parks and outdoor recreation. *Leisure Sciences, 27*, 157-173.

Legare, A. M., & Haider, W. (2008). Trend analysis of motivation-based clusters at the Chilkoot Trail National Historic site of Canada. *Leisure Sciences, 30*, 158-176.

Loker-Murphy, L. (1996). Backpackers in Australia: A motivation-based segmentation study. *Journal of Travel & Tourism Marketing, 5*, 23-45.

Manfredo, M. J. (1984). The comparability of onsite and offsite measures of recreation needs. *Journal of Leisure Research, 16*, 245-249.

Manfredo, M. J., & Larson, R. A. (1993). Managing for wildlife viewing recreation experiences: An application in Colorado. *Wildlife Society Bulletin, 21*, 226-236.

Manfredo, M. J., Driver, B. L., & Tarrant, M. A. (1996). Measuring leisure motivation: A meta-analysis of the recreation experience preference scales. *Journal of Leisure Research, 28*, 188-213.

Manning, R. E. (1999). *Studies in outdoor recreation: Search and research for satisfaction*. Corvallis, OR: Oregon State University Press.

Manning, R. E. (2007). *Parks and carrying capacity: commons without tragedy.* Washington DC: Island Press.

Manning, R., Lawson, S., Newman, P., Laven, D., Valliere, W. (2002). Methodological issues in measuring crowding-related norms in outdoor recreation. *Leisure Sciences, 24*, 339-348.

May, J. A., Bastain, C. T., Taylor, D. T., & Whipple, G. D. (2001). Market segmentation of Wyoming snowmobilers. *Journal of Travel Research, 29*, 292-299.

McLaughlin, W. J. & Paradice, W. E. J. (1980). Using visitor preference information to guide dispersed winter recreation management for cross-country skiing and snowmobiling. In *Proceedings: North American symposium of dispersed winter recreation.* St. Paul, MN: Office of Special Programs, Agricultural Extension Service, University of Minnesota, pp. 64-72.

Milligan, G. W. & Cooper, M. C. (1986). A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate Behavioral Research, 21*, 441-458.

Morey, L. C., Blashfield, R. K., & Skinner, H. A. (1983). A comparison of cluster analysis techniques within a sequential validation framework. *Multivariate Behavioral Research, 18*, 309-329.

Morgan, G. A., Leech, N. L., Glockner, G. W., & Barret, K. C. (In collaboration with Clay, J. N., Jenson, L. & Quick, D.). (2004). *SPSS for introductory statistics: use and interpretation* (2nd ed.). Mahway, NJ: Lawrence Earlbaum Associates.

Mosteller, F., & Tukey, J. W. (1968). Data analysis, including statistics. In Lindzey, G. & Aronson, E. (Eds.), *Handbook of social psychology* (Volume 2). Reading, MA: Addision-Wesley.

Needham, M. D., Vaske, J. J., Donnelly, M. P., Manfredo, M. J. (2007). Hunting specialization and its relationship to participation in response to chronic wasting disease. *Journal of Leisure Research, 39*, 413-437.

Nyaupane, G. P., White, D. D., & Budruk, M. (2006). Motive-based tourist market segmentation: an application to Native American cultural heritage sites in Arizona, USA. *Journal of Heritage Tourism, 1*, 81-99.

Oh, C., Ditton, R. B., Anderson, D. K., Scott, D., & Stoll, J. R. (2005). Understanding differences in nonmarket valuation by angler specialization level. *Leisure Sciences, 27*, 263-277.

Petrick, J. F., Backman, S. J., Bixler, R., & Norman, W. C. (2001). Analysis of golfer motivations and constraints by experience use history. *Journal of Leisure Research, 33*, 56-70.

Park, H., & Jun, C. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications, 36*, 3336-3341.

Salter, D. W. (2003). A generalizability study of Salter Environmental Type Assessment scores by social climate domain. *Measurement and Evaluation in Counseling and Development, 36*, 130-139.

Schomaker, J. H., & Knopf, R. C. (1982). Generalizability of a measure of visitor satisfaction with outdoor recreation. *Applied Psychological Measurement, 6*, 173-183.

Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist, 44*, 922-932.

Shinew, K. J., Floyd, M. F., & Parry, D. (2004). Understanding the relationship between race and leisure activities and constraints: exploring an alternative framework. *Leisure science, 26*, 181-199.

Shores, K., & Scott, D. (2007). The relationship of individual time perspective and recreation experience preferences. *Journal of Leisure Research, 39*, 28-59.

Stewart, S. C., Fix, P. J., & Manfredo, M. J. (2004). P*ublic perceptions of elk and vegetation management in Rocky Mountain National Park, Colorado.* (Project Rep. No. 56). Project Rep. for Rocky Mountain National Park. Fort Collins: Colorado State University, Human Dimensions in Natural Resources Unit.

Strube, M. J. (2000). Psychometrics. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G. (Eds.), *Handbook of psychophysiology.* Cambridge, UK: Cambridge University Press.

Stuhlmann, J., Daniel, C., Dellinger, A., Denny, R. K., & Powers, T. (1999). A generalizability study of the effects of training on teachers' abilities to rate children's writing using a rubric. *Journal of Reading Psychology, 20*, 107-127.

Swanson, J. E., Vande Kamp, M. E., & Johnson, D. R. (2002). Trip motivations. In Swanson, J. E., Vande Kamp, M. E., Johnson, D. R., Manning, R. E., and Lawson, S.R. (Eds.) *A survey of overnight backcountry visitors to Denali National Park and Preserve.* USGS Technical Report NPS/CCSOUW/NRTR-2002-04, Seattle, WA: Forest and Rangeland Ecosystem Science Center.

Vaske, J. J. (2008). *Survey research and analysis: applications in recreation, parks, and human dimensions.* State College, PA: Venture Publishing.

Vaske, J. J., & Donnelly, M. P. (2002). Generalizing the encounter – norm – crowding relationship. *Leisure Sciences, 24*, 255-269.

Vaske, J. J., Donnelly, M. P., Heberlein, T. A., & Shelby, B. (1982). Differences in reported satisfaction ratings by consumptive and nonconsumptive recreationists. *Journal of Leisure Research, 14*, 195-206.

Vaske, J. J., Donnelly, M. P., Wittmann, K., & Laidlaw, S. (1995). Interpersonal versus social values conflict. *Leisure Sciences, 17*, 205-222.

Vaske, J. J., & Manning, R. E. (2008). Analysis of multiple data sets in outdoor recreation research: introduction to the special issue. Leisure Sciences, 30, 93-95.

Vaske, J. J., Needham, M. D., & Cline Jr., R. C. (2007). Clarifying interpersonal and social values conflict among recreationists. *Journal of Leisure Research, 39*, 182-195.

Vaske. J. J., & Shelby, L. B. (2008). Crowding as a descriptive indicator and evaluative standard: results from 30 years of research. *Leisure Sciences, 30*, 111-126.

Vaske, J. J., Whittaker, D., Shelby, B., & Manfredo, M. J. (2002). Indicators and standards: developing definitions of quality. In Manfredo, M .J., Vaske, J. J., Bruyere, B. L., Field, D. L., & Brown, P. J. (Eds.), *Society and natural resources: A summary of knowledge*. Jefferson, MO: Modern Litho. Pgs. 143-171.

Walker, G. J., Jackson, E. L., & Deng, J. (2008). The role of self-construal as an intervening variable between culture and leisure constraints: a comparison of Canadian and Mainland Chinese university students. *Journal of Leisure Research, 48*, 90-109.

White, D. D., Virden, R. J., & Cahill, K. L. (2005). Visitor experiences in National Park Service cultural sites in Arizona: implications for interpretive planning and management. *Journal of Park and Recreation Administration, 23*, 83-97.
Williams, D. R., Ellis, G. D., Nickerson, N. P., & Shafer, C. S. (1988). Contributions of time, format, and subject to variation in recreation experience preference measurement. *Journal of Leisure Research, 20*, 57-68.
Williams, D. R. & Vaske, J. J. (2003). The measurement of place attachment: validity and generalizability of a psychometric approach. *Forest Science, 49*, 830-840.

Footnotes

[1] Stability is also used to compare consistency in scores of the same rater across time. While we expect our results to be similar, there were different raters in the time periods we were comparing.

[2] Breckenridge (2000) compares clusters derived from a secondary cluster analysis with both samples combined to the original cluster assignments, whereas Morey, Blashfield, and Skinner (1983) compare the results of two partitions of the data.

[3] Generalizability Theory uses the term facet rather than factors to avoid confusion with factor analysis.

[4] Chi-square and independent samples t tests were used to test the following variables by five categories of visitation (6/11 to 6/21, 6/22 to 7/10, 7/11 to 7/30, 7/31 to 8/18, 8/19 to 9/05) and visitation before and on/after July 8: Alaska residency, length of stay, method of travel, group membership and size, primary activity, primary source of information, number of children in group, age and the motivation domains included in the study. The only significant difference found was for the exercise domain (p =.042, n = 228), with means of 3.5 and 3.8 for before July 8 and July 8 or after, respectively. However, the effect size (Hedges G) was .28, indicating a minimal relationship, and an inspection of the data also suggests the difference is not of practical significance.

[5] The mill town is spelled as Kennecott, whereas the glacier and valley are spelled as Kennicott.

[6] One reason for refusals was that visitors were in a rush to catch their shuttle, which left approximately every half hour and the shuttles, when parked, were visible from inside the mill town (hence many visitors would stay in the mill town until the last possible minute). However, most of those refusing did have time to answer the two nonresponse questions. The p values for the nonresponse tests for children present, time of day, prior visitation, and gender, as tested with Chi-square, and number in group and structure stabilization impacting trip, tested with a t test, are as follows for t_1 and t_2: .15/.81, .08/.71, .76/na, .81/.88, .21/.75, .24/.89.