

Inter-Observer Agreement of a Multi-Parameter Campsite Monitoring Program on the Dixie National Forest, Utah

Nicholas J. Glidden
Martha E. Lee

Abstract—Precision is crucial to campsite monitoring programs. Yet, little empirical research has ever been published on the level of precision of this type of monitoring programs. The purpose of this study was to evaluate the level of agreement between observers of campsite impacts using a multi-parameter campsite monitoring program. Thirteen trained observers assessed 16 dispersed campsites on the Dixie National Forest in southern Utah. The data were analyzed using Cohen's kappa statistic to determine the level of observer agreement beyond chance for each of the impact attributes and the condition class rating. Results showed a moderate level of proportional agreement and a low level of agreement beyond chance. These results suggest that the data collection protocol must be improved to increase the level of inter-observer agreement. Results also indicate that managers should evaluate their training procedures to increase precision.

Introduction

Outdoor recreation in the United States has increased dramatically during the past 50 years (Cole 1999). This increase is related to a number of factors, including a larger population, greater mobility, more leisure time, greater affluence, improved recreational technology, and a rise in private vehicle ownership (Cole 1999; Sun and Walsh 1998). As outdoor recreation grows, adverse ecological impacts increase as well (Cole 1999; Cole and others 1996; McEwen and Cole 1997; Williams and Marion 1995). Managers' concern for the effects of recreation on natural resources led to the emergence of the field of recreation ecology.

Recreation ecology is the field of study that examines, assesses, and monitors visitor impacts (Leung and Marion 2000). Early antecedents date back to E. P. Meinecke's (1928) examination of tourist impacts on the root system of redwood trees in California, and G. H. Bates's (1935) study of trampling effects on vegetation adjacent to footpaths in England. In the 1960s, recreation ecology began to expand as the number of studies grew and the first scientific conferences were held on the subject (Cole 1999). A considerable body

of information on recreation ecology has built up since that time, with contributions from scientists around the world (Cole and Schreiner 1980).

As the field of recreation ecology continues to grow, its application to land management has become more recognizable. During the past two decades, most wildland managers became aware of and concerned about biophysical impacts but had little objective information about impact levels or trends over time upon which to manage impacts (Cole 2000). A lack of objective impact information commonly led to whimsical and inconsistent impact management plans (Cole 2000).

Public acceptance of ineffective management plans diminished as natural resource management issues became increasingly controversial (Cole 2000). The shift in public scrutiny and participation in the planning process, combined with the passing of the National Environmental Policy Act of 1969 (42 USC 4321 *et seq.*), forced recreation managers to facilitate goal-oriented planning systems, such as the Limits of Acceptable Change (LAC) and Visitor Experience and Resource Protection (VERP) frameworks (Cole 2000; Stankey and others 1985). These frameworks rely on systematic and objective recreation impact assessments and monitoring data on which management decisions are based (Cole 2000). As a result, many recreation managers have developed ecological impact monitoring programs.

Monitoring programs vary in methods used to collect data (Leung and Marion 2000). The choice of methods is based on the questions asked, types of data needed for management, character of the study area, training of investigators, and logistical constraints (Leung and Marion 2000). Due to activity concentration and duration of stay, campsites receive high levels of visitor impacts, and are of primary concern in recreation impact monitoring (Cole 1994, 1995; Leung and Marion 2000; Williams and Marion 1995).

Campsite monitoring programs consist of a primary assessment process, often referred to as an inventory, where site impact information is assessed on a number of impact attributes on a sample or census of sites followed by a reinventory process where site impact attributes are reassessed at a determined interval. Most campsite monitoring programs include a 5-year reinventory cycle (Cole 1989b).

Management of wildland recreation areas requires valid research and monitoring data for preparing, assessing, and implementing effective policy (Cole 2000; Williams and Marion 1995). Campsite monitoring programs, when appropriately designed and implemented, can supply answers to a broad range of management information needs (Leung and Marion 2000). Effective monitoring programs provide managers with longitudinal data to detect changes

Nicholas J. Glidden, Recreation Specialist, Dixie National Forest, UT, U.S.A.

Martha E. Lee, Professor of Recreation Management, School of Forestry, University of Northern Arizona, Flagstaff, U.S.A.

In: Watson, Alan; Sproull, Janet; Dean, Liese, comps. 2007. Science and stewardship to protect and sustain wilderness values: eighth World Wilderness Congress symposium: September 30–October 6, 2005; Anchorage, AK. Proceedings RMRS-P-49. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

in ecological impact from recreation use, as well as evaluate and make informed adjustments on existing management plans (Cole 1989a, 2000; Landres 1995; Leung and Marion 2000; Williams and Marion 1995). Monitoring programs help managers anticipate trends in recreation impacts rather than react to impacts, providing opportunity to prevent the proliferation of site impacts by implementing sound management plans (Cole 1990, 2000).

Campsite impact assessment and monitoring methods include photographic assessments, condition class approaches, quantitative measurements of multiple parameters, as well as combinations of all three methods (Leung and Marion 2000). Photographic systems were among the first applied and are based on repeat photographs taken from permanently established photo points (Williams and Marion 1995). The benefits of the photographic system are that it's relatively quick and easy. However, few impact types can be accurately evaluated using photographs and it is difficult to assign interval level ratings for site comparison (Cole 1989a). Condition class systems are based on observers comparing site conditions to pre-determined descriptive condition classes, and recording the class that most closely matches the conditions (Williams and Marion 1995). This system is also quick and easy, but observer bias tends to be high and lacks quantifiable data on specific resource conditions (Cole 1989b; Williams and Marion 1995). Multi-parameter systems are based on individual measurements and appraisals of specific indicators of resource condition. These systems require greater observer training and often take longer to perform but can yield more accurate and precise measurements of campsite conditions (Leung and Marion 2000; Williams and Marion 1995).

Over the last decade, emphasis has been placed on improving campsite monitoring procedures to increase the level of precision (Cole 1989a; Leung and Marion 2000; Williams and Marion 1995). Imprecise data may make it difficult for managers to distinguish real change over time from separate imprecise estimates of the same impact (Cole 1989a). It is also particularly relevant in monitoring programs where multiple individuals will be collecting information on the same sites at different occasions over time. In other words, the temporal and longitudinal nature of the monitoring process makes precision especially important.

The precision of a monitoring program refers to the variability in estimates of campsite impact indicators by many observers (Williams and Marion 1995). The level of precision is also referred to as the random error of the study. High precision indicates that random variation associated with the collection procedure is minimized (Province of British Columbia 1998). Precision can be studied through statistical analysis of repeated measurements. Unfortunately, because one individual or a group of individuals typically assess a campsite, at one point in time there is only one set of data or one measurement of an attribute, making it impossible to calculate the level of precision. Techniques that yield precise data are particularly important to government agencies where turnover of monitoring personnel is high. If the number of campsites in a monitoring program is large enough that it takes multiple field seasons, and different monitoring personnel are hired each season (as is typical with federal and state agencies), then it is important for the observers to be consistent in evaluating the impact attributes. Therefore, inter-observer agreement should be

a priority in any monitoring program. Knowing the level of precision of a monitoring program enables managers to identify the difference between two dissimilar measures of the same condition and a real change in the condition over time (Cole 1989b).

The precision of a monitoring program is crucial to its fundamental purpose, yet precision has never been empirically evaluated. It is the intent of this study to empirically evaluate the inter-observer agreement of the assessment process of a multi-parameter campsite monitoring program.

Methods

Study Site

Area Description. The study site is located on the Dixie National Forest in southern Utah. This is the largest national forest in Utah, occupying almost 2 million acres, and stretching for about 170 miles (274 km) east to west, straddling the divide between the Great Basin and the Colorado River. The forest is comprised of diverse ecosystems. Forest elevations range from 2,800 to 11,322 feet (853 to 3,451 m) above sea level, with annual precipitation ranging from 10 to 40 inches (25 to 102 cm), and temperatures ranging from -30 to over 100 degrees Fahrenheit (-34 to over 38 degrees Celsius). Vegetation types grade from sparse, sagebrush (*Artemisia tridentata*) and rabbitbrush (*Chrysothamnus nauseosus*) at lower elevations, low-growing pinyon pine (*Pinus edulis*) and juniper (*Juniperus scopulorum*) at mid-elevations, to aspen (*Populus tremuloides*), ponderosa pine (*Pinus ponderosa*), bristlecone pine (*Pinus longaeva*), Englemann spruce (*Picea engelmannii*), and subalpine fir (*Abies lasiocarpa*) at high elevations. This variety of environmental settings provides for a variety of recreation activities on the forest.

Recreation activities include hiking alpine mountains and slot canyons, mountain biking slickrock and single-track trails, and riding all terrain vehicles along the Great Western Trail. However, camping is one of the primary recreation activities in the forest. A recent inventory of campsites found over 1,800 dispersed campsites and over 80 wilderness campsites (Glidden 2001, 2002, data on file with the Dixie National Forest). The Dixie National Forest receives a great deal of recreation use by visitors to the three surrounding national parks (Zion, Bryce, and Capitol Reef) and one national monument (Grand Staircase-Escalante). In 2003, this national forest had approximately 2.5 million visitors (Max Molyneux, personal conversation on March 16, 2004).

Campsites. A sample of previously inventoried dispersed campsites (non-wilderness and non-developed) were used in this study. The sites were selected using a multi-staged stratified sampling approach. The forest was stratified into two areas delineated by political ranger district boundaries, the Cedar City Ranger District and the Panguitch Ranger District. The two districts contained 427 and 343 dispersed campsites, respectively, for a total of 770 sites. These two strata were then divided into sub-unit plots delineated by the Universal Transverse Mercator (UTM) coordinate system. Two sub-unit plots in each district were randomly selected using a random number generator. Within each of the two randomly selected sub-unit plots, four campsites were randomly selected for assessment. Therefore, 16 sites

(four sites in each of the four groups) were assessed. This method allowed for random selection of sites while maintaining a high level of efficiency by allowing a representative sub-sample of sites to be assessed.

Monitoring Program

The monitoring program used to evaluate level of precision was the Dixie National Forest Campsite Monitoring Program (DCM). This monitoring program was developed in the summer of 2002 for dispersed campsite inventories, and combines the use of photographic, condition class, and multi-parameter techniques adapted from Cole (1989b) and Marion (1995). The program is unique in that it incorporates Global Positioning System (GPS) and Geographic Information System (GIS) technology to obtain and utilize data. The DCM includes 12 site characteristics and 12 site impact attributes (table 1). Because only the impact indicators are used in the overall impact index, I focused my study on these data along with the condition class rating.

Monitoring Procedure

The first step of the monitoring process is to identify the site. For this study, the sites were selected and marked with a site identification number prior to data collection. Once a site is identified, its boundary is determined and recorded using a GPS. The DCM attribute impact data are then assessed and entered into a GPS unit via a data dictionary. The data are then downloaded from the GPS unit and differentially

corrected at the office. The tabular data, or observed data, were exported into a pre-formatted *Excel* spreadsheet that calculates the impact index totals.

Observer Characteristics

The 13 observers who participated in this research were Dixie National Forest employees at the time of the project. Observers varied in age, gender, employment status (seasonal or fulltime), and job title. The observers also varied in experience and educational backgrounds (table 2). This group of individuals represented the range of experience and educational backgrounds common among seasonal monitoring personnel.

Training and Data Collection

Individuals were given the DCM manual one week prior to the training to familiarize themselves with the protocol, followed by a day of training. Training began with time for the observers to re-read the manual and review the manual with the instructor, taking time to answer questions, and then proceeding to a demonstration site to systematically review the procedures specified by the protocol. After forming groups of three or four, each group evaluated the same four sites, reviewed with the instructor the sites, and discussed questions that arose. Observers were then divided into four different groups and evaluated another four sites. The idea behind forming new groups was to expose and ameliorate possible biases by comparing data from the different groups.

Table 1—Table of campsite characteristics and impact indicators and their associated attribute choices.

Site characteristic	Attribute choices
Date	(The date the site was inventoried)
Time	(The time the site was inventoried)
Photo/Site ID number	(Identification number based on the number of the first photograph taken of the site)
Inventoried by	(Initials of the person surveying the site)
Type of site	(Campsite) (Stock Tie Site) (Other)
Site location	(Meadow) (Forested area) (Rock) (Lakeside) (Streamside)
Site access	(2-WD) (4-WD) (ATV) (Hiking/Equestrian)
Potential site expansion	(Poor) (Moderate) (Good)
Number of campsite w/in sight	(Manually entered number of sites within sight)
Firewood availability	(Yes) (No)
Barren core camp area	(% of site that 90% or more of vegetation is absent)
Frissell condition class rating	(1) (2) (3) (4) (5)
Site impact attribute	Attribute choices
Vegetation cover onsite	(0-5%) (6-25%) (26-50%) (51-75%) (76-100%)
Vegetation cover offsite	(0-5%) (6-25%) (26-50%) (51-75%) (76-100%)
Mineral soil exposure onsite	(0-5%) (6-25%) (26-50%) (51-75%) (76-100%)
Mineral soil exposure offsite	(0-5%) (6-25%) (26-50%) (51-75%) (76-100%)
Tree damage	(0-5%) (6-25%) (26-50%) (51-75%) (76-100%)
Root exposure	(0-25%) (25-50%) (51-75%) (76-100%)
Presence of noxious weeds	(Yes) (No)
Evidence of stock	(None) (Feed / Manure) (Manure odor / Dishing)
Amount of development	(None) (Primitive structure) (Temporary structure) (Permanent structure)
Cleanliness	(No trash or manure) (Trash present) (Manure present) (Trash and manure present)
Presence of human waste	(Yes) (No)
Number of social trails	(None) (1-2) (3 or more)

Table 2—Experience and educational background of study observers.

Experiential/Educational background	Number of observers
	(n = 13)
Classes in recreation	5
Classes in forestry/ecology/wildlife	7
High school education	2
Post high school education	6
Completed undergraduate degree	3
Post undergraduate education or graduate degree	2
Previous resource impact assessment training/courses	4
Previously performed campsite assessments	5

Once the group site evaluations were complete, the sites were reviewed once again.

The 2-day data collection period began with the assessment of eight sites on the Cedar City Ranger District followed by the eight sites on the Powell Ranger District. Each observer went into each of the sites alone, assessed the campsite impact attributes, and entered the data into the DCM data dictionary on a GPS unit. Discussion among the observers was prohibited during the data collection process.

Data Analysis

To assess inter-observer agreement on campsite impact attributes, data collected were analyzed using *Gkappa version 2.3* statistical software package developed in 1993 by Dr. John Uebersax to determine a kappa value. Kappa is a widely utilized statistical tool used to assess the level of agreement among raters, ultimately evaluating the reliability of an established protocol. This method has been used in a number of disciplines, including psychiatry, epidemiology, and wildlife biology (Lantz and Nebenzahl 1996; Neuman and others 1999; Rockwood and others 2000). For example, Neuman and others (1999), used kappa to evaluate inter-observer agreement between individuals observing nesting habits of birds. The kappa value is calculated based on two factors, proportion of observed agreement and proportion of chance agreement (Cohen 1960; Uebersax 1983). In this way, kappa is considered to be chance adjusted, taking into account the level of agreement among observers that may occur by chance alone.

The two most common versions of kappa are those developed by Cohen (1960) and Fleiss (1971). Cohen's original kappa is used to evaluate two raters, but later an extension was added to accommodate multiple raters (Komagata 2002). Cohen (1968) introduced a weighted kappa that allows for the seriousness of disagreement to be accounted for. For example, for attributes with ordinal variables such as categories of proportion mineral soil onsite (0 to 5 percent, 6 to 25 percent, 26 to 50 percent, 51 to 75 percent, and 76 to 100 percent), the researcher may want to consider the severity of the disagreement (Goodwin 2001). In this case, disagreement as to whether the vegetation cover onsite is 0 to 5 percent or 76 to 100 percent is more severe than disagreement between 0 to 5 percent and 6 to 25 percent. Though Cohen's weighted kappa sounds attractive, it requires arbitrary weights be assigned to the rating categories prior to data collection

and is not recommended by some researchers (MacLure and Willett 1987; Tinsley and Brown 2000). In addition, since some of the observers did not use all the rating categories and one of the rating categories was not used at all, software capable of calculating weighted kappa for the data of this study was not available. Weighted kappa was therefore not used to calculate kappa in this research.

Fleiss' kappa, introduced in 1971, was developed to evaluate the inter-rater agreement between multiple raters. This version of kappa differs from Cohen's kappa in the way in which the proportion of chance agreement is calculated (Komagata 2002). Komagata (2002) argues that Fleiss' kappa is less desirable than Cohen's kappa due to the way each formula accounts for chance. Cohen's kappa computes chance based on the individual rater's judgment, whereas, Fleiss' kappa computes chance by averaging out the probability of all categories of the raters, placing a larger assumption on the even distribution of responses between the categories (Komagata 2002). Due to the categorical nature of this study and the emphasis on observational data, the data were not equally distributed among categories. Thus, Fleiss' kappa was not chosen for this study.

The Cohen's unweighted kappa statistic was used because the data were from multiple observers, categorical (nominal and ordinal), not normally distributed, and responses were not equally distributed among rating categories. The kappa statistic is based on a rating from -1 to 1, with 1 being 100 percent agreement beyond chance. For most purposes, kappa values from -1 to 0.4 represent a low level of agreement beyond chance, 0.4 to 0.75 represents a fair to moderate level of observer, and 0.75 to 1 represent a high or excellent level of agreement beyond chance (Banerjee and others 1999). The inter-observer agreement between each observer was calculated for each site attribute for all sites. Some of the data collected were subject to the prevalence and bias effects associated with kappa. These paradoxes are associated with the uneven distribution of ratings between attribute choices (Hoehler 1999). For example, if observers choose the 0 to 5 percent root exposure category for most of the sites assessed, there would then be an uneven distribution of response variables across the rating categories. When present, the prevalence and bias effects tend to decrease kappa values (Hoehler 1999).

In addition to calculating the kappa value, the proportion of observer agreement was also calculated. The proportion of agreement was reported to supplement the kappa statistic to address the effects of the prevalence and bias effects aforementioned. It should be noted that the proportion of

agreement is not adjusted for chance agreement. Therefore, the proportion of agreement values should only be extrapolated to the observers in the study, whereas the kappa values may more accurately represent a wider range of monitoring personnel using this style of program (Goodwin 2001).

Results

I quantified the level of inter-observer agreement of 13 observers for the 12 impact attributes presented in table 1 and found a moderate level of proportional agreement among observers, and a low level of agreement beyond chance (kappa). The proportion of observer agreement for each impact attribute ranged from 91 percent for root exposure to 35 percent for tree damage (fig. 1). The campsite impact attributes that showed the highest level of agreement (75 percent to 100 percent) among observers were: amount of development, evidence of stock, human waste, and root exposure. The campsite impact attributes that showed a moderate level of agreement (40 percent to 75 percent) were cleanliness, difference in mineral soil on and offsite, difference in vegetation cover on and offsite, mineral soil exposure offsite, mineral soil exposure onsite, number of social trails, vegetation cover offsite, and vegetation cover onsite. The lowest level of agreement (0 percent to 40 percent) among observers was for tree damage.

Kappa statistics for the impact attributes ranged from a high of 0.569 for the presence of human waste to a low of 0.155 for tree damage (fig. 2). Campsite impact attributes that showed a moderate level of observer agreement beyond chance (0.4 to 0.75) were evidence of stock, and human waste. The campsite impact attributes that showed the lowest levels of observer agreement beyond chance were cleanliness, development, difference in mineral soil exposure, difference in vegetation cover, mineral soil exposure offsite, mineral soil exposure onsite, root exposure, number of social trail, tree damage, vegetation cover offsite, and vegetation cover onsite.

The difference between the proportion of agreement and the kappa value for each attribute was substantial, with kappa values always being lower than the proportion of agreement (fig. 3). We would expect the kappa values to be lower because kappa accounts for the chance of observer agreement rather than just the level of agreement.

Overall, certain attributes showed consistent levels of inter-observer agreement from both the kappa statistic and the proportion of agreement. The difference in vegetation cover, mineral soil exposure onsite, tree damage, and vegetation cover onsite consistently showed low (< 0.5) kappa and proportion of agreement values, whereas evidence of stock and presence of human waste consistently showed moderate to high (>0.5) kappa and proportion of agreement values.

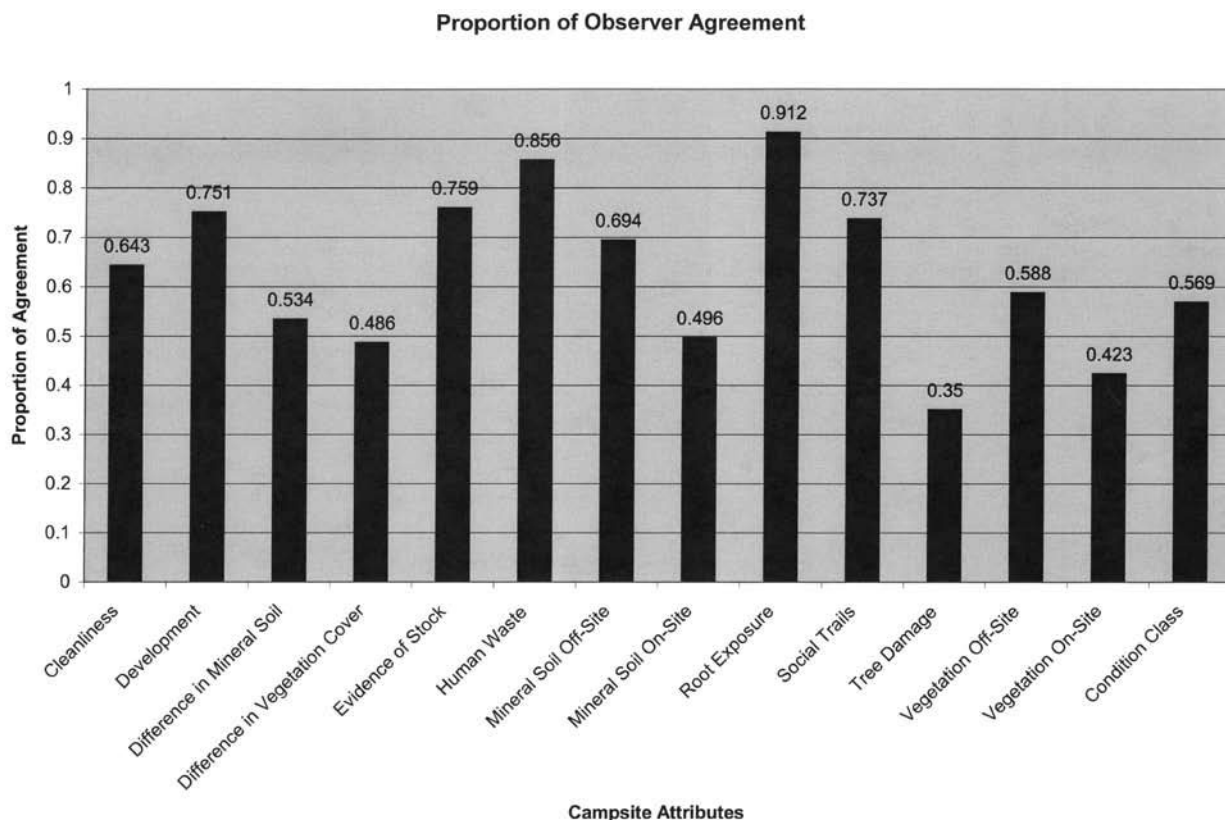


Figure 1—Percent of observer agreement for each impact attribute across all sites.

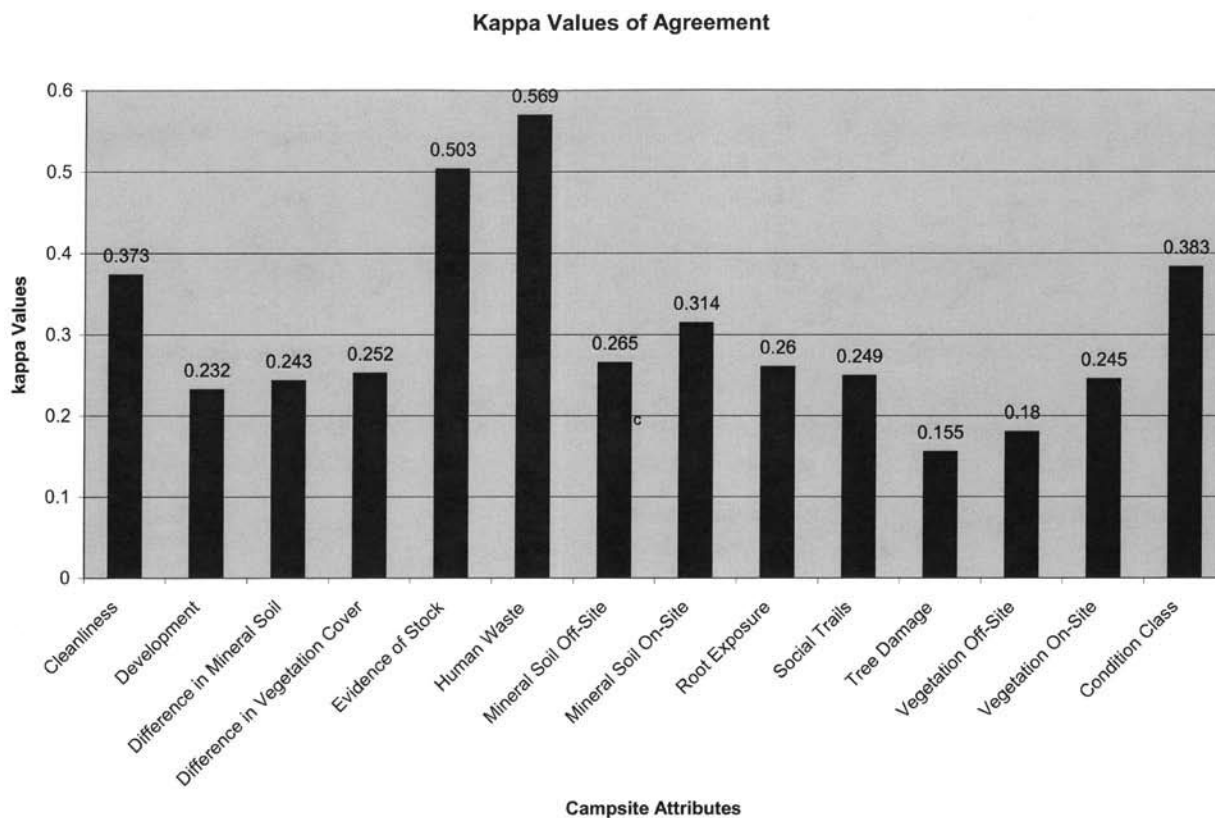


Figure 2—Level of observer agreement beyond chance (kappa) for each impact attribute.

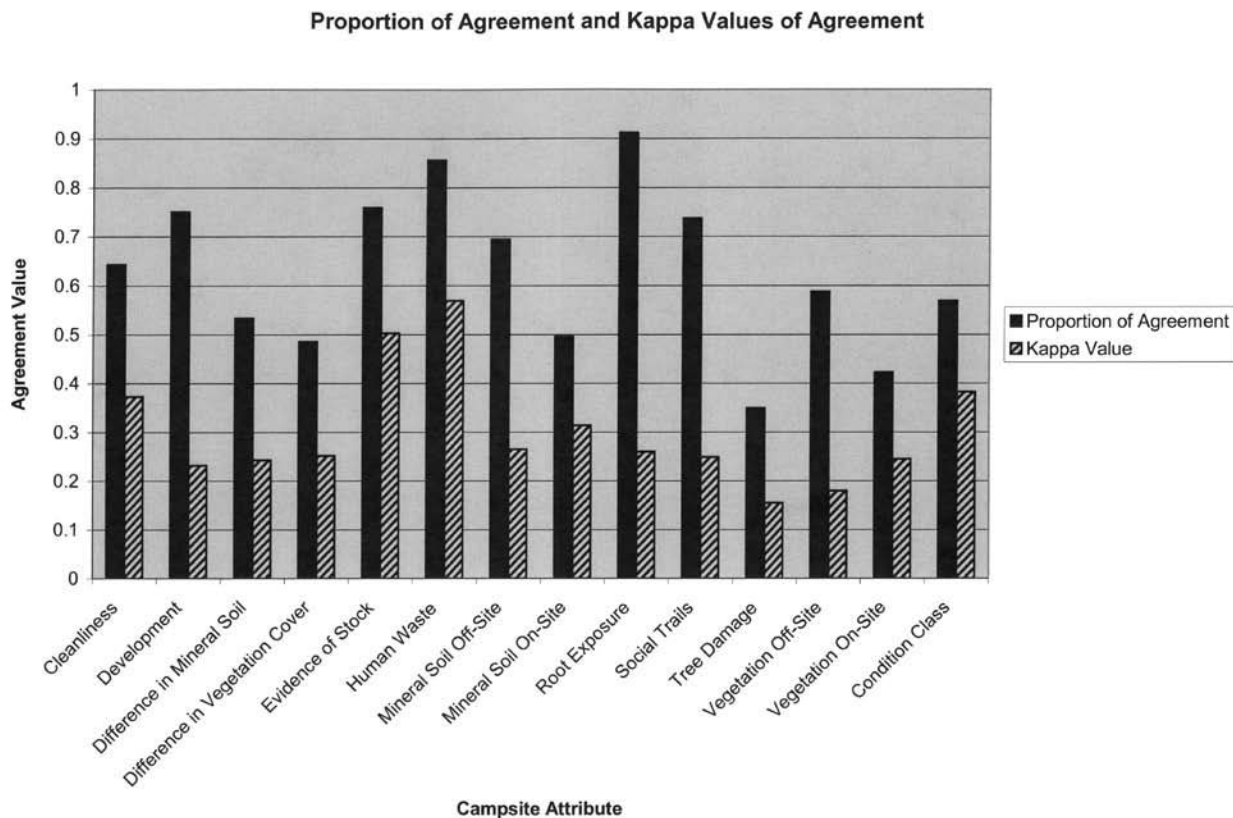


Figure 3—Comparison of agreement between proportion of agreement and kappa analysis.

Management Implications

The results of this study have important implications to recreation managers and researchers, especially for agencies that hire seasonal monitoring personnel. The moderate to low levels of precision suggest that managers using this or similar multi-parameter monitoring programs should thoroughly evaluate the protocol for collecting the impact indicators to identify any possible sources for observer confusion and/or bias.

The results indicate the protocol for collecting the percent mineral soil (on and offsite), the percent vegetation cover (on and offsite), and the amount of tree damage need to be improved to increase precision. In particular, these results indicate the precision of attributes based on whole-site measurements, such as percent vegetation cover, rather than a specific feature of the site, such as level of development, should be assessed. Impact indicators that consistently yield imprecise data may need to be given less weight in the overall campsite impact index, or removed from the monitoring program entirely.

Managers and researchers should also improve training procedures to increase the level of precision. Although the effect of the amount and complexity of observer training has never been empirically tested, it makes intuitive sense that properly trained individuals are likely to report more precise data. Training should be focused on improving the level of precision of impact attributes that yield imprecise data. The results of this research suggest additional training is needed to improve the level of precision of the percent vegetation cover, percent mineral soil exposure, and the amount of tree damage. One way this could be accomplished is by performing test site inventories and having observers calibrate off of one another. This process would involve a comparison of the observers' attribute rating, and a subsequent mental adjustment by each observer that would make the reported data more consistent with each other. Another way to improve the level of precision through training is by having an inexperienced observer work with an experienced observer, or just by having two observers work together. Once again, this will allow the observers to calibrate off of one another and also alleviate some biases.

The final implication of this research is that managers should calculate the level of precision of their campsite impact monitoring program and determine how much change would need to occur in the monitoring data to be considered a true change in campsite condition. By knowing the level of precision, managers will be able to decipher whether a true change has occurred or whether two observations of the same level of impact have yielded two different reported levels of impact. That is, managers can assume that the estimated trend in campsite condition will be more meaningful as the precision of the protocol increases. This concept becomes more pertinent as the number of sites per management decision decreases.

Conclusions

Precision is particularly important to the fundamental purpose of a campsite monitoring program of predicting trends in site impacts. This study has shown that a moderate to low level of inter-observer agreement exists in the multi-parameter campsite monitoring program used on the

Dixie National Forest. The kappa values may be lower due to the prevalence and bias effect, because of the unevenly distributed ratings between attribute choices. For example, observers tended to rate root exposure in the 0 to 5 percent category for all sites, so the level of proportional agreement is high but the kappa value is low. It should be noted that kappa values associated with attributes subject to the prevalence and bias effect, such as root exposure, number of social trails, and the amount of development, should be viewed with caution.

The results of this study are particularly important for managers using this type of campsite monitoring program. Because this is the first study of the level of inter-observer agreement of a campsite monitoring program, additional research is needed to evaluate the level of precision of other types of campsite monitoring programs.

The fundamental question of this research has application beyond the field of recreation management. Due to the high rate of monitoring personnel turnover, the level of precision should be evaluated for all ecological monitoring programs. It is my hope that this research will act as an impetus for managers to evaluate the level of precision of their monitoring programs, which will result in a more effective administration of recreation resources.

References

- Banerjee, Mousumi; Capozzoli, Michelle; McSweeney, Laura. 1999. Beyond kappa: a review of interrater agreement measures. *The Canadian Journal of Statistics*. 27(1): 3–23.
- Bates, G. H. 1935. The vegetation of footpaths, sidewalks, cattracks and gateways. *Journal of Ecology*. 23: 468–487.
- Province of British Columbia Ministry of Environment, Lands and Parks. 1998. Species inventory fundamentals: standards for components of British Columbia's biodiversity. Ministry of Environment, Lands and Parks, Resources Inventory Branch for the Terrestrial Ecosystems Task Force, Resource Inventory Committee. No. 1, Version 2.0.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 20: 27–36.
- Cohen, J. 1968. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*. 70: 213–220.
- Cole, David N. 1989(a). Area of vegetation loss: a new index of campsite impact. Research Paper. INT-389. Ogden, UT: U.S. Department of Agriculture, Forest Service, Inter-Mountain Research Station. 51 p.
- Cole, David N. 1989(b). Wilderness campsite monitoring methods: a sourcebook. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. Gen. Tech. Rep. INT-259. 57 p.
- Cole, David N. 1990. Some principles to guide wilderness campsite management. In: Lime, David W., ed. *Managing America's enduring wilderness resource: proceedings of the conference*, Minneapolis, Minnesota, September 11–17, 1989. St. Paul, Minn.: Tourism Center, Minnesota Extension Service and Minnesota Agricultural Experiment Station, University of Minnesota: 181–187.
- Cole, David N. 1994. Backcountry impact management: lessons from research. *Trends*. 31(3): 10–14.
- Cole, David N. 1995. Disturbance of natural vegetation by camping: Experimental applications of low-level stress. *Environmental Management*. 19(3): 405–416.
- Cole, David N. 1999. Recreation, ecological impacts. In: Alexander, David E., Fairbridge, Rhodes W., eds. 1999. *Encyclopedia of environmental science*. Dordrecht, The Netherlands: Kluwer Academic Publishers (sold and distributed in North America by Kluwer Academic Publishers, Hingham, MA): 506–508.

- Cole, David N. 2000. Biophysical impacts of wildland recreation use. In: Gartner, W. C.; Lime, D. W., eds. Trends in outdoor recreation, leisure and tourism. (Ch 23). New York, NY: CABI Publishing: 257–264.
- Cole, David N.; Schreiner, Edward G. S., compilers. 1980. Impacts of backcountry recreation: site management and rehabilitation—an annotated bibliography. Gen. Tech. Rep. INT-121. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experimental Station. 58 p.
- Cole, David N.; Knight, Richard L. 1990. Impacts of recreation on biodiversity in wilderness. In: Wilderness areas: their impacts; proceedings of a symposium; 1990 April 19–20; Logan, UT. Logan, UT: Utah State University: 33–40.
- Cole, David N.; Hall, Troy E. 1992. Trends in campsite condition: Eagle Cap Wilderness, Bob Marshall Wilderness, and Grand Canyon National Park. Res. Pap. INT-453. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 40 p.
- Cole, David N., Watson, Alan E., Hall, Troy E., Spildie, David R. 1996. High-use destinations in wilderness: social and biophysical impacts, visitor responses, and management options. INT-RP-496. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 30 p.
- Fleiss, Joseph L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 76(5): 378–382.
- Goodwin, Laura D. 2001. Interrater agreement and reliability. Measurement in physical education and exercise science. 5(1): 13–34.
- Hoehler, Fred K. 1999. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *Journal of Clinical Epidemiology*. 53(2000): 499–503.
- Komagata, Nobo. 2002. Chance agreement and significance of the kappa statistic. [Online]. Available: www.tcnj.edu/%7Ekomagata/pub/kappa.pdf. [May 3, 2004].
- Landres, Peter B. 1995. The role of ecological monitoring in managing wilderness. *TRENDS/Wilderness Research*. 32(1): 10–13.
- Lantz, Charles; Nebenzahl, Elliott. 1996. Behavior and interpretation of the K statistic: resolution of the two paradoxes. *Journal of Clinical Epidemiology*. 49(4): 431–434.
- Leung, Yu-Fai, Marion, Jeffrey L. 2000. Recreation impacts and management in wilderness: a state-of-knowledge review. In: Cole, David N.; McCool, Stephen F.; Borrie, William T.; O'Loughlin, Jennifer, comps. 2000. Wilderness science in a time of change conference—Volume 5: Wilderness ecosystems, threats, and management; 1999 May 23–27; Missoula, MT. Proceedings RMRS-P-15-Vol-5. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 23–48.
- Maclure, M.; Willett, W. C. 1987. Misinterpretation and misuse of the kappa statistic. *American Journal of Epidemiology*. 126: 161–169.
- Marion, Jeffrey L. 1995. Environmental Auditing: Capabilities and Management Utility of Recreation Impact Monitoring Programs. *Environmental Management*. 19: 763–771.
- McEwen, Douglas; Cole, David N. 1997. Campsite Impact In Wilderness Areas. *Parks & Recreation*. 32(2): 24–30.
- Meinecke, Emilio. 1928. A report on the effect of excessive tourist travel on the California redwood parks. Sacramento, CA: California State Printing Office. 20 p.
- Neuman, J.; Chardine, J. W.; Perter, J. M. 1999. Approaches to testing inter-observer reliability of field-collected behavioral data. *Waterbirds*. 22(3): 348–357.
- Rockwood, Kenneth; Strang, David; MacKnight, Chris; Downer, Robert; Morris, John C. 2000. Interrater reliability of the clinical dementia rating in a multicenter trial. *Journal of the American Geriatrics Society*. 48(5): 558–559.
- Stankey, G. H., Cole, D. N., Lucas, R. C., Peterson, M. E., Frissell, S. S. 1985. Limits of acceptable change (LAC) system for wilderness planning. General Technical Report. INT-176. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 37 p.
- Sun, D., Walsh, D. 1998. Review of studies on environmental impacts of recreation and tourism in Australia. *Journal of Environmental Management*. 53: 323–338.
- Tinsley, Howard E. A.; Weiss, David J. Interrater reliability and agreement. 2000. In: Tinsley, Howard E. A.; Brown, Steven D., eds. Handbook of applied multivariate statistics and mathematical modeling. San Diego, CA: Academic Press: 95–124.
- Uebersax, John S. 1983. A design-independent method for measuring the reliability of psychiatric diagnosis. *Journal of Psychiatric Research*. 17(4): 335–342.
- Williams, Peter B., Marion, Jeffrey L. 1995. Assessing campsite conditions for limits of acceptable change management in Shenandoah National Park. Technical Report NPS/MARSHEN/NRTR-95/071. Philadelphia, PA: National Park Service Chesapeake System Support Office Resource Stewardship and Partnerships. 138 p.