# An assessment of interrater reliability from the System for Observing Play and Recreation in Communities (SOPARC) in urban parks in New York City

***Myron F. Floyd**, Department of Parks, Recreation, and Tourism Management, North Carolina State University, American. mffloyd@ncsu.edu*
***J. Aaron Hipp**, Department of Parks, Recreation, and Tourism Management, North Carolina State University, American.*
***Oriol Marquet**, Department of Parks, Recreation, and Tourism Management, North Carolina State University, Spanish.*
***Claudia Alberico**, Department of Parks, Recreation, and Tourism Management, North Carolina State University, Brazilian.*
***Elizabeth Mazak**, Department of Parks, Recreation, and Tourism Management, North Carolina State University, American.*
***Jing Huang**, Department of Parks, Recreation, and Tourism Management, North Carolina State University, Chinese.*

---

SOPARC reliability is affected by features of the target area and by contextual conditions present at the time of observation.

*Observations made in playgrounds, during the weekend, or in areas with consistently high user counts, had lower reliability scores.*

## Introduction

Systematic observational studies in open environments such as parks and playgrounds can be challenging because of the lack of objective measures to quantify visitor characteristics and behavior. Since its introduction in 2006, the SOPARC protocol (Systematic Observation of Play and Recreation in Communities) (McKenzie et al. 2006) has provided an answer to those challenges. SOPARC is based on momentary sampling through periodic and systematic scans of pre-delimited target areas. It provides a consistent method to count large groups of people while they are taking part in often highly dynamic activities without placing a burden on participants (Cohen et al. 2011).

After a period of training, reported overall interrater reliability at counting people in parks using SOPARC has been consistently high (Banda et al. 2014; Chung-Do et al. 2011). However, there is a lack of understanding of how SOPARC reliability might be affected by the factors such as time of day, presence of organized activities or type of target areas, could affect reliability and agreement between observers. This variance might be exacerbated when trying to record combined user attributes such as age and gender or age and physical activity, measurements that have traditionally drawn lower reliability scores (Floyd et al. 2011; Chung-Do et al. 2011; Floyd et al. 2008).

To address this gap, we examine the extent interrater reliability varies across various types of target areas and contextual conditions using 3390 paired SOPARC observations in New York City parks.

## Methods

### *Data gathering*

SOPARC observations were taken in 20 New York City parks within a larger study of Physical Activity and Recreation of Children in Communities of Color (PARC[3]). The PARC[3] research team selected a total of 167 target areas within the parks. These included playgrounds, swings and water features, sports courts, and fields. Trained observers used a modified SOPARC to characterize park users according to gender, age group, ethnicity, and phyiscal activity. Each target area was observed four times between April and June 2017 and four times between July and August 2017. Visits were repeated twice for each weekend day during each season and observations were hour-long taken at 10:00, 15:00, 16:30, and 18:00. A total of 4725 observations were taken in pairs. From those observations and following McKenzie's et al. (2006) method, those with at least one person counted by at least one of the raters were analyzed (n=3390).

### *Measures of reliability*

Based on the 3390 paired observations we computed the mean of counts of park user stratified by age and gender. Intraclass Correlation (ICC) was used to determine the reliability between the two observers. A variant of the overall proportion of agreement between observers was also computed. In areas with at least one observer reporting more than 10 people, a 10% discrepancy was allowed and coded as agreement (McKenzie et al. 2006).
Both reliability indicators were analyzed in diverse observation (day of week, time of day, and round of SOPARC), target area (e.g., playgrounds, sport courts, etc.) and diverse contextual conditions (target areas, extent of formal and organized activities, extent of shade cover).

### *Statistical analysis*

Mean counts by gender and age and ICC scores are reported in Table 1. Additionally, we regressed the odds of two observers agreeing on the number of people present in a target area, under each different observation condition. Binary logistic regression models were fitted to estimate the effect of observation conditions on the observers' agreement. Marginal means representing the chance of the two observers agreeing on the number of people in the target area while adjusting for all the contextual variables of the target area are reported to facilitate interpretation (Table 1).

## Results

General interrater reliability for the 3390 paired observations was excellent (ICC=0.941). The difficulty of trying to assess gender and age at the same time however decreased reliability of the observations to a range between 0.81 (male 5 to 10 years) and 0.91 (female teenagers). Reliability consistently decreased when observing target areas early in the day (10-11am), target areas where an informally organized activity was taking place, and when observing basketball courts. On the other hand, reliability was consistently better in the afternoon (6-7pm), and in completely shaded areas (except for 5 to 10 and teenage boys).
Analyzing the adjusted marginal means on agreement between observers shows observations made during weekends were significantly less reliable at counting children of both genders from 0 to 10 years old. Having informally organized activities in the target area also was significantly associated with higher reliability when counting both boys and girls between 5 and 10 years old. Counting teenage boys in informally organized activities on the other hand scored lower reliability than in areas without organized activities.

In terms of types of target areas, playgrounds were found to be the areas where reaching agreement was difficult. Reliability when counting children for both genders and between 0 and 10 years old was significantly higher in all areas that were not playgrounds, except for basketball courts in the case of males age 5-10. For teenagers, reliability was significantly lower in swing areas in the case of girls, and in basketball courts in the case of boys. On the other hand, counting girls in basketball courts and teenagers in "other areas" was associated with higher reliability scores.

**Table 1**: Reliability measures for observation of children in distinct target area conditions.

| | | 0 to 4 | | | | | | 5 to 10 | | | | | | Teenagers | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Female | | | Male | | | Female | | | Male | | | Female | | | Male | | |
| | N of scans | Cpo | ICC | Agree[1] | Cpo | ICC | Agree[1] | Cpo | ICC | Agree[1] | Cpo | ICC | Agree[1] | Cpo | ICC | Agree[1] | Cpo | ICC | Agree[1] |
| Day of week | | | | | | | | | | | | | | | | | | | |
| Weekday | 2528 | 0.48 | 0.86 | 0.89 | 0.43 | 0.84 | 0.89 | 0.83 | 0.84 | 0.84 | 1.24 | 0.78 | 0.78 | 0.49 | 0.93 | 0.91 | 1.48 | 0.85 | 0.82 |
| Weekend | 862 | 0.70 | 0.84 | 0.84* | 0.77 | 0.85 | 0.83* | 1.25 | 0.83 | 0.77* | 1.72 | 0.86 | 0.73* | 0.50 | 0.84 | 0.92 | 1.48 | 0.91 | 0.81 |
| Time of day | | | | | | | | | | | | | | | | | | | |
| 10-11am | 239 | 0.38 | 0.43 | 0.77 | 0.40 | 0.65 | 0.78 | 0.55 | 0.36 | 0.73 | 0.98 | 0.23 | 0.67 | 0.12 | 0.57 | 0.92 | 0.52 | 0.37 | 0.83 |
| 3pm-4pm | 1011 | 0.47 | 0.87 | 0.90* | 0.47 | 0.82 | 0.89* | 0.90 | 0.78 | 0.85* | 1.35 | 0.84 | 0.79* | 0.44 | 0.93 | 0.93 | 1.36 | 0.90 | 0.85 |
| 4.30-5.30pm | 668 | 0.53 | 0.84 | 0.87* | 0.45 | 0.87 | 0.88* | 0.78 | 0.80 | 0.82* | 1.19 | 0.81 | 0.75* | 0.58 | 0.92 | 0.89 | 1.76 | 0.82 | 0.81 |
| 6-7pm | 1471 | 0.61 | 0.89 | 0.89* | 0.61 | 0.87 | 0.87* | 1.10 | 0.91 | 0.82* | 1.51 | 0.88 | 0.77* | 0.55 | 0.89 | 0.91 | 1.59 | 0.89 | 0.80 |
| Round | | | | | | | | | | | | | | | | | | | |
| 1 | 1011 | 0.54 | 0.87 | 0.88 | 0.51 | 0.85 | 0.86 | 0.89 | 0.86 | 0.82 | 1.40 | 0.82 | 0.77 | 0.50 | 0.93 | 0.91 | 1.55 | 0.90 | 0.80 |
| 2 | 995 | 0.52 | 0.85 | 0.88 | 0.50 | 0.85 | 0.88 | 0.92 | 0.80 | 0.82 | 1.34 | 0.74 | 0.75 | 0.55 | 0.89 | 0.91 | 1.53 | 0.86 | 0.83 |
| 3 | 875 | 0.51 | 0.87 | 0.89 | 0.53 | 0.86 | 0.87 | 0.92 | 0.84 | 0.84 | 1.30 | 0.85 | 0.77 | 0.49 | 0.90 | 0.90 | 1.46 | 0.88 | 0.83 |
| 4 | 509 | 0.60 | 0.84 | 0.87 | 0.56 | 0.83 | 0.88 | 1.10 | 0.87 | 0.82 | 1.45 | 0.86 | 0.75 | 0.36 | 0.91 | 0.93 | 1.29 | 0.76 | 0.82 |
| Formally organized | | | | 0.88 | | | 0.87 | | | | | | | | | | | | |
| N | 3346 | 0.54 | 0.86 | - | 0.53 | 0.85 | - | 0.95 | 0.84 | 0.82 | 1.36 | 0.81 | 0.76 | 0.43 | 0.86 | 0.91 | 1.49 | 0.87 | 0.82 |
| Y | 65 | 0.08 | - | - | 0.03 | - | - | 0.24 | - | - | 2.31 | 0.64 | - | 3.68 | 0.97 | 0.74* | 1.58 | 0.76 | 0.78 |
| Informally organ. | | | | | | | | | | | | | | | | | | | |
| N | 2803 | 0.65 | 0.85 | 0.88 | 0.63 | 0.84 | 0.87 | 1.12 | 0.85 | 0.82 | 1.42 | 0.81 | 0.76 | 0.49 | 0.92 | 0.91 | 0.91 | 0.84 | 0.85 |
| Y | 877 | 0.02 | - | - | 0.01 | - | - | 0.11 | 0.30 | 0.90* | 1.28 | 0.77 | 0.80* | 0.45 | 0.78 | 0.90 | 4.19 | 0.80 | 0.76* |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Shade | | | | | | | | | | | | | | | | | | | |
| Sun | 1354 | 0.63 | 0.82 | 0.86 | 0.58 | 0.82 | 0.87 | 1.07 | 0.85 | 0.81 | 1.35 | 0.82 | 0.75 | 0.50 | 0.92 | 0.90 | 1.24 | 0.83 | 0.82 |
| Mix | 1653 | 0.50 | 0.86 | 0.90* | 0.49 | 0.84 | 0.87 | 0.93 | 0.78 | 0.84* | 1.30 | 0.79 | 0.79 | 0.53 | 0.90 | 0.92 | 1.51 | 0.91 | 0.83 |
| Shade | 1143 | 0.44 | 0.87 | 0.88 | 0.43 | 0.87 | 0.88 | 0.79 | 0.87 | 0.81 | 1.44 | 0.76 | 0.74 | 0.45 | 0.93 | 0.92 | 1.84 | 0.84 | 0.80 |
| Facility type | | | | | | | | | | | | | | | | | | | |
| Playground | 809 | 1.06 | 0.81 | 0.79 | 1.10 | 0.82 | 0.78 | 1.79 | 0.83 | 0.72 | 1.81 | 0.85 | 0.70 | 0.31 | 0.81 | 0.90 | 0.16 | 0.73 | 0.94 |
| Basketball | 379 | 0.00 | - | - | 0.01 | - | - | 0.09 | 0.74 | 0.96* | 1.39 | 0.69 | 0.74 | 0.36 | 0.87 | 0.96* | 4.48 | 0.78 | 0.59* |
| Baseball/Handball | 1120 | 0.03 | 0.96 | - | 0.03 | 0.83 | - | 0.12 | 0.61 | 0.92* | 0.49 | 0.81 | 0.88* | 0.83 | 0.93 | 0.93 | 1.73 | 0.81 | 0.82* |
| Swings | 508 | 0.82 | 0.85 | 0.86* | 0.59 | 0.80 | 0.88* | 1.34 | 0.86 | 0.78* | 0.92 | 0.83 | 0.85* | 0.82 | 0.92 | 0.83* | 0.33 | 0.87 | 0.93 |
| Water | 237 | 0.50 | 0.91 | 0.92* | 0.56 | 0.87 | 0.87* | 0.95 | 0.84 | 0.85* | 1.16 | 0.86 | 0.79* | 0.26 | 0.85 | 0.92 | 0.20 | 0.80 | 0.93 |
| Other | 337 | 0.24 | 0.78 | 0.94* | 0.23 | 0.72 | 0.93* | 0.48 | 0.41 | 0.89* | 1.61 | 0.82 | 0.77* | 0.69 | 0.93 | 0.92 | 1.01 | 0.86 | 0.83* |
| TOTAL | 3390 | 0.54 | 0.86 | 0.88 | 0.52 | 0.85 | 0.87 | 0.94 | 0.84 | 0.82 | 1.36 | 0.81 | 0.76 | 0.49 | 0.91 | 0.91 | 1.48 | 0.86 | |

Cpo. = average counts per observation

1. Marginal means from binary logistic regressions. Relaxed rule agreement between observers (10% rule)

* Significantly different from reference category. P<0.05. Reference category is always the first category of the nominal variable.

- Insufficient sample size (less than 20 observations per cell)

## Discussion

The SOPARC protocol continues to show high reliability across diverse target areas and contexts. Paired observers reached consistent agreement across all categories of gender and age. Similarly, consistent excellent agreement (ICC>0.75) was achieved throughout the majority of target area types and observation conditions. The findings demonstrated that specific attention should be given to observations of large numbers of people, as all observations counting an average of more than 3 people per target area had significantly lower reliability. Other factors such as time of the day and types of target areas should also receive greater attention during training of SOPARC observers. Particularly, observation of playgrounds, with their often-intricate designs and consistently high number of children, should be improved to ensure better reliability.

## References

Banda, J.A. et al., 2014. The associations between park environments and park use in southern US communities. *Journal of Rural Health*, 30(4), pp.369–378.

Chung-Do, J.J. et al., 2011. An observational study of physical activity in parks in Asian and Pacific Islander communities in urban Honolulu, Hawaii, 2009. *Preventing chronic disease*, 8(5), p.A107. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3181180&tool=pmcentrez&rendertype=abstract.

Cohen, D.A. et al., 2011. How much observation is enough? Refining the administration of SOPARC. *Journal of physical activity & health*, 8(8), pp.1117–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22039130%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3689589.

Floyd, M.F. et al., 2011. Park-based physical activity among children and adolescents. *American Journal of Preventive Medicine*, 41(3), pp.258–265. Available at: http://dx.doi.org/10.1016/j.amepre.2011.04.013.

Floyd, M.F. et al., 2008. Park-Based Physical Activity in Diverse Communities of Two U.S. Cities. An Observational Study. *American Journal of Preventive Medicine*, 34(4), pp.299–305.

McKenzie, T.L. et al., 2006. System for Observing Play and Recreation in Communities (SOPARC): Reliability and Feasibility Measures. *Journal of Physical Activity and Health*, 3(s1), pp.S208–S222. Available at: http://journals.humankinetics.com/doi/10.1123/jpah.3.s1.s208.