J. Field Ornithol., 70(4):481–490

CAN NOVICE VOLUNTEERS PROVIDE CREDIBLE DATA FOR BIRD SURVEYS REQUIRING SONG IDENTIFICATION?

MARGARET A. MCLAREN
Ontario Ministry of Natural Resources
R.R. 2 Hwy 11 North
Bracebridge, Ontario, P1L 1W9 Canada

MICHAEL D. CADMAN
Canadian Wildlife Service
75 Farquhar St.
Guelph, Ontario, N1H 3N4 Canada

Abstract.—We tested whether interested volunteers with low to moderate skill levels could be trained to identify and count a subset of forest birds, by song or call, well enough to provide credible data. Paired novice and experienced observers performed five simultaneous, unlimited distance 10-min point counts of 12 (1995) or 13 (1996) target species. We found no differences in either counts of individual species or in the suite of species present between experienced-experienced pairs and experienced-novice pairs. Nevertheless, novices tended to count fewer birds of most target species than experienced observers. These differences were significant for Red-eyed Vireo, Rose-breasted Grosbeak and Scarlet Tanager in 1995 but only for Rose-breasted Grosbeak in 1996. Of the nine novices who participated in both years, all improved their pre-training ability to identify the target species (from a tape) between 1995 and 1996. The counts of Red-eyed Vireos (the only species with a sample size large enough to test) were significantly lower than the counts of the experienced observers in 1995 but not in 1996. We conclude that trained and tested novice observers can provide credible data for a carefully selected subset of species.

¿PUEDEN LOS NOVICIOS PROVEER DATOS CONFIABLES EN CENSOS DE AVES EN DONDE SE UTILIZA EL CANTO COMO FUENTE DE IDENTIFICACIÓN?

Sinopsis.—Tratamos de probar si voluntarios con destrezas bajas y moderadas sobre la identificación de aves, podían ser adiestrados para identificar aves de bosques utilizando sus llamadas y cantos, y producir entonces datos confiables. Parejas de novicios y observadores experimentados llevaron a cabo, simultáneamente, un censo de 10 minutos (sin límite de distancia) de 12 (1995) y 13 (1996) especies particulares. Los acuerdos a que se llegaron en el número de especies y número de individuos (por especie) entre las parejas fueron puestos a comprobación utilizando una prueba Wilcoxon. En el 1996 también se puso a pruebas los acuerdos de identificación entre los observadores experimentados. No se encontraron diferencias en la tasa de acuerdos en el contaje de especies o de individuos entre los individuos pareados o entre los observadores experimentados. Sin embargo, los novicios tendieron a contar menos aves de la gran mayoría de las especies seleccionadas que los observadores experimentados. Estas diferencias resultaron ser significativas para *Vireo olivaceus, Pheucticus ludovicianus* y *Piranga olivacea* en el 1995, pero tan solo para *P. ludovicianus* en 1996. Los nueve novicios en el experimento, mejoraron en su habilidad para identificar aves entre 1995 y 1996. Los conteos del vireo (la única muestra grande que se pudo analizar estadísticamente) resultaron significativamente menor que los que contaron los observadores experimentados en el 1995, pero no así para el 1996. Se concluye que los novicios que son adiestrados adecuadamente pueden proveer datos confiables para una serie de especies cuidadósamente seleccionadas.

In recent years, growing concern about the status of a variety of songbird species has led both governments and private organizations to place increasing emphasis on gathering information on which to assess popu-

lation status over broad areas. The North American Breeding Bird Survey (BBS) (Robbins et al. 1986) has been a major source of data for analyses of broad-scale trends. However, the BBS is a road-side survey and its ability to provide adequate information about species that inhabit only interior forests, particularly those with quiet songs, is an open question. In Ontario, the Forest Bird Monitoring Program (FBMP), a point count-based survey co-ordinated by the Canadian Wildlife Service (CWS), was implemented in 1987 to supplement information from the BBS, particularly for interior forest species.

As a result of a legal review of forest management planning in Ontario, the Ontario Ministry of Natural Resources (OMNR), is required to assess and report on the population status of representative wildlife species potentially affected by forest management on public lands in the central and northern parts of the province. The area to be assessed consists of over 500,000 km^2 with a total human population of about 1.3 million. The number of potential volunteers with the necessary skills to undertake an FBMP survey is low.

Obtaining an adequate number of skilled observers is often a limiting factor in the scope of volunteer-based monitoring projects. The availability of volunteers has been an important factor in determining the size and distribution of blocks to be covered in breeding bird atlas projects. When necessary, "priority" blocks are included to help ensure that specified minimum coverage requirements are met (e.g., Erskine 1992). To expand the number of potential volunteers, the Marsh Monitoring Program (Anonymous 1996), provides volunteers with a training tape cassette including the vocalizations of target species and instructions on how to identify them and differentiate them from similar-sounding species.

Despite the low numbers of skilled volunteers in northern Ontario, our preliminary investigations suggested that there are many people in northern parts of the province who would like to participate in surveys leading to better conservation and management of birds. The need for better information on bird populations over this large and sparsely populated area led us to investigate the ability of interested but low-skilled volunteers to increase their skills to a level at which they could provide reliable data for a subset of forest species using the FBMP protocol. We call the protocol using a subset of species the Streamlined Forest Bird Monitoring Program (SFBMP).

METHODS

Novice observers.—Novices willing to participate in the test were located through solicitation via the OMNR electronic bulletin board and contacts with naturalist clubs. The study co-ordinator discussed the protocol with novices and then mailed each novice a training package. All novices self-assessed their pre-training ability to identify birds by sound as poor to moderate. The study co-ordinator also matched the novice with an experienced FBMP participant whose survey site was within reasonable driving distance (about 50 km) for the novice.

Training package.—The training package consisted of an introductory letter explaining the program, a set of instructions identical to that received by the regular (experienced) FBMP participants, an audio tape, a written description of the songs and call notes of the target species, a set of data forms, a sheet for recording answers to a quiz, and a comment sheet. Side A of the audio tape consisted of an identification quiz, which used the same recordings, but in random order and without voice identification of the species, as the training side of the tape (side B). Side B consisted of sound clips of the target species in taxonomic order, with each clip introduced by the name of the species. Recordings used in training were largely, but not entirely, made in Ontario and included both call notes and territorial songs (or drumming pattern in the case of Yellow-bellied Sapsucker). In 1995, the identification quiz consisted only of the target species. In response to comments from participants that inclusion of additional species would make the quiz more realistic (i.e., more like the situation in the field), the quiz in 1996 included additional nontarget species. Novices were told that there were species they were not expected to recognize in the quiz. The training tape continued to consist of only the target species.

Survey protocol.—Novices were asked to listen to Side A of the audio tape and to write the names of species they recognized on the answer sheet provided in their training package before either listening to side B of the tape or reading the written descriptions of the songs and calls of the target species. After doing the quiz they were to train themselves using Side B and the written instructions. They were encouraged but not required to do additional field training either alone or with an experienced observer. After training, they were instructed to repeat the quiz without reference to their initial answers.

Survey sites in the regular FBMP consist of five permanent counting stations that are at least 100 m from a forest edge and 250 m apart (Welsh 1995). Two 10-min point counts (the first between 25 May and 17 June (Visit 1) and the second between 13 June and 7 July (Visit 2), with at least 6 d between counts) are conducted at each station. During analysis, the higher of the two counts (totalled over the five stations) for each species is used. Many sites are surveyed by a single person. Assistants may accompany the surveyor but they are not permitted to point out birds that may have been missed. FBMP data are collected by volunteers who have been assessed (based either on personal acquaintance or through telephone discussion) as sufficiently skilled by Cadman.

For the SFBMP test, we matched novices with FBMP participants who had contributed at least one year of data to the program. Novices were asked to accompany the FBMP observer on one or the other of his/her field visits. Both FBMP and SFBMP observers were instructed that there should be no discussion of either species present or numbers of individuals during the counts. The two observers independently recorded numbers of each species at each station on the field sheets provided and each independently transferred data from field sheets to summary sheets at a

later time. FBMP observers counted all birds they detected at each station, whereas novices recorded only the SFBMP target species. SFBMP observers submitted quiz results (both before and after training), field and summary sheets, and any comments on the program to CWS. FBMP observers also submitted their field and summary sheets to CWS. For both SFBMP and FBMP observers, the transfer of data from field to summary sheets was checked for accuracy. The matched visit was then identified and the data for the SFBMP target species were extracted from the FBMP summary sheet for that visit. The values compared in the analyses presented here were the number of species and the total number of individuals of each species recorded at a site. In 1996, we also tested the agreement in number of species and number of individuals (by species) detected between pairs of experienced observers. Four experienced observers surveyed the five stations at an established FBMP site simultaneously, yielding six pair-wise comparisons.

Target species.—Species were selected for testing using four criteria: sufficiently abundant that they would be recorded frequently enough to allow for statistical testing, having songs that we considered would not be too confusing for novices (e.g., we did not include species such as Chipping Sparrow [*Spizella passerina*] which we considered could be easily confused with Dark-eyed Junco [*Junco hyemalis*], having relevance to conservation issues (e.g., neotropical migrants believed to be in decline), and appropriate to the three forest regions of Ontario.

We used three overlapping sets of species (Table 1) to account for differences in bird distribution across Ontario. From south to north, road-accessible Ontario spans 10 degrees of latitude and three forest zones (deciduous, Great Lakes-St. Lawrence, and boreal) (Rowe 1972). Numerous species reach the northern or southern limits of their geographical range in the province necessitating differing lists of target species. Consideration of 1995 results led to some adjustments in the species lists in 1996. Specifically, we added American Robin to test the possibility that novices were confusing this species with other similar sounding species, we removed a few rarely recorded species, and we re-aligned the geographic lists so that there was more species overlap. The latter was done in an attempt to increase the sample size of usable pairs of observers. Testable species were those that were recorded by at least one member of six novice-experienced pairs. Six is the smallest sample for which the Wilcoxon matched-pairs signed-ranks test can produce statistically significant results at $P < 0.05$.

Statistical tests.—We compared the rate of agreement on number of individuals (by species) between experienced-experienced pairs and experienced-novice pairs using a chi square contingency table of agree vs. disagree. For rate of agreement on species presence, we used a 3×2 contingency table of agree present, agree absent, or disagree. Comparisons between experienced and novice observers of numbers of birds of each species detected were made using the Wilcoxon matched-pairs signed-ranks test. $P < 0.05$ was considered significant even though we

conducted multiple tests. Because of the conservation implications of our study, we preferred to accept the null hypothesis erroneously (Type II error) than to reject it erroneously.

RESULTS

In 1995, we distributed 82 training kits and received 24 returns, which yielded 18 usable pairs for comparison. In 1996, we distributed 89 training kits and received 38 returns, which yielded 33 usable pairs. Reasons for unusable returns included failure of the two observers to do their counts on the same date (although they were at the same site), failure of the regular FBMP volunteer to return data, and discussion of the birds by the two observers during the count. Despite efforts to recruit and match novices from central and northern Ontario, fewer than 20 matches were made in each of these areas. Return rate of data was similar from the three areas of the province but the small number of matches in central and northern areas resulted in a substantial number of species for which the sample size was too small ($n < 6$) for statistical testing.

Novice volunteers who did submit data varied greatly in their initial skill levels. In 1995, before-training score for the 16 novices who submitted quiz results ranged from 2.5–12 (mean = 7.4) out of 12. In 1996, before-training scores for 29 novices ranged from 0–13 (mean = 9.7) out of 13 but this was influenced by the inclusion of nine people who participated in both years. These nine people increased their mean before-training score from 8.4 out of 12 to 12.3 out of 13 between the two years. The mean score of new participants in 1996 was 7.5 (range: 0–13). In 1995, all of the novices identified the 12 target species correctly after training. In 1996, the mean score after training was 12 (range: 10–13).

Variability between experienced-experienced pairs vs. experienced-novice pairs was tested using 1996 data only. No differences were found in overall agreement on counts of number of individuals of a species ($\chi^2 = 0.47$, df = 1, $P > 0.05$) or detection of individual species ($\chi^2 = 5.6$, df = 2, $P > 0.05$).

Novice observers did tend to count fewer individuals than experienced observers (Table 2). Differences were significant for Red-eyed Vireo, Rose-breasted Grosbeak, and Scarlet Tanager in 1995 but only for Rose-breasted Grosbeak in 1996. We suspect that this result was influenced by the increased skill level of the nine observers who had participated in 1995, but we could test only Red-eyed Vireo for observers who participated in both years. For this species, novices counted significantly fewer individuals in 1995 (Wilcoxon matched-pairs signed-ranks test, $P < 0.02$, $n = 8$) but not 1996 (Wilcoxon matched-pairs signed-ranks test, $P > 0.05$, $n = 7$).

We added American Robin as a target species in 1996 because of concerns that novices were mis-identifying other species with robin-like songs (Red-eyed Vireo, Scarlet Tanager, Rose-breasted Grosbeak) as robins. There was no evidence to suggest that this was the case. Rather, novices seemed to be simply not counting robin-like songs when they were uncertain of the true identity.

TABLE 1. Target species used to test the identification and counting skills of novice observers in 1995 and 1996. Testable species were those for which a statistically significant result was possible using a Wilcoxon matched-pairs signed-ranks test ($n > 5$ pairs). Data for the two years could not be pooled because 9 of 24 participants in 1996 had also contributed data in 1995. Some species are shown for only one year because the target list changed slightly between 1995 and 1996.

Species	1995		1996	
	Testable	Not testable	Testable	Not testable
Yellow-bellied Sapsucker (*Sphyapicus varius*)				X
Eastern Wood-Pewee (*Contopus virens*)	X		X	
Least Flycatcher (*Empidonax minimus*)		X		X
Great Crested Flycatcher (*Myiarchus crinitus*)	X		X	
Red-eyed Vireo (*Vireo olivaceus*)	X		X	
Black-capped Chickadee (*Poecile atricapillus*)	X		X	
Red-breasted Nuthatch (*Sitta canadensis*)		X		X
Winter Wren (*Troglodytes roglodytes*)	X		X	
Ruby-crowned Kinglet (*Regulus calendula*)		X		X
Veery (*Catharus fuscesens*)		X	X	
Swainson's Thrush (*Catharus ustulatus*)		X		X
Hermit Thrush (*Catharus guttatus*)		X	X	
Wood Thrush (*Hylocichla mustelina*)	X		X	
American Robin (*Turdus migratorius*)		X		
Black-throated Blue Warbler (*Dendroica caerulescens*)		X		X
Black-throated Green Warbler (*Dendroica virens*)		X		X
Black-and-white Warbler (*Mniotilta varia*)				X
Ovenbird (*Seiurus aurocapillus*)	X		X	
Scarlet Tanager (*Piranga olivacea*)	X		X	

TABLE 1. Continued.

Species	1995		1996	
	Testable	Not testable	Testable	Not testable
White-throated Sparrow (*Zonotrichia albicollis*)		X		X
Rose-breasted Grosbeak (*Pheucticus ludovicianus*)	X		X	
Brown-headed Cowbird (*Molothrus ater*)		X	X	

DISCUSSION

Numerous authors have addressed the problems of observer bias and observer errors in singing bird surveys (e.g., Bystrak 1981, Robbins and Stallcup 1981, Bart and Schoultz 1984, Bart 1985). These studies demonstrate that even expert observers under-count, over-count, and misidentify birds under particular circumstances. The importance of training observers, particularly in research settings, is also well recognized (Kepler and Scott 1981, Hanowski and Niemi 1995). However, for broadly based volunteer-dependent surveys such as the BBS and FBMP, observers tend to be self-selected and screening of their bird identification skills is minimal.

For the most part, volunteers in these programs are likely conscientious and aware of the importance of accurate identification and counting (but see Faanes and Bystrak 1981). Informal contacts with some of the novices in this study who did not submit their data, despite going in the field with an experienced FBMP participant, suggest that their reasons for nonsubmission were largely related to their perception that their data were of no value because they "were not good enough." Our intent in this study was not to investigate observer bias *per se,* but rather to assess whether novices could identify, count, and record a small number of species at the same rates as experts. We assumed that effects of weather conditions, differential sound conduction in different habitats, and the known biases relating to counting difficulties when numerous birds are singing simultaneously (Bart and Schoultz 1984, Bart 1985) would apply equally to novices and experts.

Faanes and Bystrak (1981) identified BBS routes conducted by volunteers they considered to be under-qualified by comparing expected numbers of species with actual numbers reported by these volunteers. Using temporal data from sets of BBS routes where the observer changed from under-qualified to qualified, they concluded that under-qualified observers detected fewer species, were more inconsistent in number of species recorded and recorded smaller numbers of individuals, particularly when detection was by song or call of species they designated as "intermediate" or "difficult to learn". In this study, we also found that novices tended

TABLE 2. Comparison of numbers of individuals of target species recorded at each site by pairs of experienced and novice surveyors in 1995 and 1996.

Species	1995				1996			
	Mean count (experienced)	Mean count (novice)	*n*	*P*[a]	Mean count (experienced)	Mean count (novice)	*n*	*P*
Eastern Wood-Pewee	8.25	8.06	16	0.87	7.23	7.37	30	0.89
Great Crested Flycatcher	3.71	2.79	14	0.17	3.25	3.24	25	0.98
Red-eyed Vireo	12.87	9.46	18	0.004	10.66	9.09	32	0.14
Black-capped Chickadee	4.29	5.50	12	0.17	2.95	3.32	22	0.08
Winter Wren	3.00	4.17	6	0.06	3.80	3.20	15	0.34
Veery	3.35	4.53	11	0.19	5.30	4.80	13	0.86
Hermit Thrush	N/A[b]	N/A	<6	N/A	6.43	5.86	7	0.48
Wood Thrush	5.30	4.64	12	0.44	6.54	6.62	13	0.95
American Robin	N/A	N/A	N/A	N/A	4.12	4.75	28	0.93
Ovenbird	11.77	11.00	13	0.44	12.76	12.60	33	0.97
Scarlet Tanager	4.17	2.10	10	0.017	2.47	1.47	15	0.07
Rose-breasted Grosbeak	6.30	3.30	10	0.01	4.80	2.50	20	0.01
Brown-headed Cowbird	N/A	N/A	<6	N/A	3.33	1.22	9	0.08

[a] Wilcoxon matched-pairs signed-ranks test.
[b] Sample size insufficient to test or no data for year (see text).

to count fewer individuals overall, but we did not find novices to be more inconsistent than experienced observers in either species detection rates or counts of individuals. However, we have reason to question this result. There was, indeed, substantial disagreement in numbers of each species counted among the experienced observers and it is likely correct to conclude that these observers were as variable among pairs as novice-experienced pairs. However, the non-significant result of the test of species detected is dependent on detection of a single Scarlet Tanager by one of the experienced observers but not the other three. We suspect that had the sample of experienced-experienced pairs been larger, there would have been better agreement on species presence between experienced pairs than between novices and experienced surveyors.

There was considerable variability in the before-training skill levels of the novices who participated in this study. Sample sizes were not sufficient to test higher skilled novices vs. lower skilled novices but the data suggest that those with higher initial skills were more similar in counts to their experienced counterpart that novices with lower initial skills. The improvement in agreement between novices and experienced observers in 1996 vs. 1995 for those who participated in both years also suggests that novices with some initial knowledge of songs of the target species will be better able to provide credible data.

Although we recognize the dangers of employing under-qualified observers, we intend to pursue the option of recruiting novice observers in northern parts of Ontario where we have no pool of skilled observers to call on. We believe that this study shows that novices can provide credible data for a carefully selected subset of species, particularly after one year of experience. Nevertheless, we do intend to screen potential volunteers and to accept only those whose skills are adequate to identify our selected subset of species consistently. We also intend to provide considerably more training than we offered for this study, including at least one trip to the field during the breeding season during which the actual survey protocol can be practiced. We received numerous comments to the effect that novices who had practiced only with the training tape found the presence of many species singing simultaneously confusing.

ACKNOWLEDGMENTS

We thank all the volunteers, both novices and experienced FBMP participants, who assisted with this study. Linda Burr coordinated development of the training packages and matched novices with experienced volunteers. The study was supported by the Ontario Ministry of Natural Resources and the Canadian Wildlife Service, Ontario Region. This paper is a contribution of the Ontario Wildlife Assessment Program.

LITERATURE CITED

ANONYMOUS. 1996. Marsh monitoring program: training kit and instructions for surveying marsh birds, amphibians and their habitats. Long Point Bird Observatory (now Bird Studies Canada) and Environment Canada, Ontario. 40 pp.

BART, J. 1985. Causes of recording errors in singing bird surveys. Wilson Bulletin 97:161–172.

——— AND J. D. SCHOULTZ. 1984. Reliability of singing bird surveys: changes in observer efficiency with avian density. Auk 101:307–318.

BYSTRAK, D. 1981. The North American Breeding Bird Survey. Studies in Avian Biology 6: 34–41.

ERSKINE A. J. 1992. Atlas of breeding birds of the maritime provinces. Nimbus Publishing Ltd. and the Nova Scotia Museum. Halifax, Nova Scotia. 270 pp.

FAANES, C. A., AND D. BYSTRAK. 1981. The role of observer bias in the North American Breeding Bird Survey. Studies in Avian Biology 6:353–359.

HANOWSKI, J. M., AND G. J. NIEMI. 1995. Experimental design considerations for establishing an off-road habitat-specific monitoring program using point-counts. Pp. 145–150, *in* C. J. Ralph, J. R. Sauer, and S. Droege, eds. Monitoring bird populations by point counts. U.S. Dept. Agr., Gen. Tech. Rep. PSW-GTR-149.

KEPLER, C. B., AND J. M. SCOTT. 1981. Reducing count variability by training observers. Studies in Avian Biology 6:366–371.

ROBBINS, C. S., AND R. W. STALLCUP. 1981. Problems in separating species with similar habitats and vocalizations. Studies in Avian Biology 6:360–365.

———, D. BYSTRAK, AND P. H. GEISSLER. 1986. The Breeding Bird Survey: its first fifteen years, 1965–1979. Resource Publication 157, U.S. Dept. Interior. Fish and Wildlife Service. Washington, D.C.

ROWE, J. S. 1972. Forest regions of Canada. Bulletin 123. Canada Forestry Service.

WELSH, D. A. 1995. An overview of the Ontario Forest Bird Monitoring Program in Canada. Pp. 93–97, *in* C. J. Ralph, J. R. Sauer, and S. Droege, eds. Monitoring bird populations by point counts. U.S. Dept. Agr. Gen. Tech. Rep. PSW-GTR-149.

Received 23 Apr. 1998; accepted 15 Dec. 1998.