Studies in Avian Biology No. 29:182–186

EFFECTIVENESS OF INFORMAL BANDING TRAINING AT THREE WESTERN CANADIAN BANDING STATIONS

Brenda C. Dale

Abstract. Skills of trainee banders from three western Canadian banding stations were assessed in 1993, using minimum performance standards in use at that time. Each trainer–trainee combination independently examined the same birds. Quantitative skills appeared to be learned quickly, but there were few passing scores on aging, sexing, skulling, and fat assessment, and none on overall achievement using the test standards. However, many trainee errors were of a non-critical nature, which was not well reflected in the scoring system. Several individuals did score well if the nature of their errors was taken into account. Time spent with a trainer, experience, and personality may all play a role in trainee performance. Results demonstrate the need for trainers to meet an established standard, and for continued spot-checking of skills after training has been completed.

Key Words: banding techniques, banding test standards, bander training.

Use of data collected at banding stations for such important international programs as Monitoring Avian Productivity and Survival (MAPS) and migration monitoring is predicated on the belief that data are collected accurately. However, standards for obtaining banding permits vary greatly throughout the world. In a few cases, a formal test is administered, but in North America permits are awarded on the basis of letters of recommendation from banders who already have permits.

Recently, the North American Banding Council (NABC) developed detailed guides on banding techniques, a guide for trainers, and specialized manuals for the banding of landbirds, hummingbirds, and raptors (Hull et al. 2001; North American Banding Council 2001a, b, c; Russell et al. 2001). Intensive training courses are increasingly available, and a bander can now undergo testing to earn formal certification. In the last few years the Canadian Bird Banding Office and the U.S. Bird Banding Lab have begun to accept certification as proof of sufficient skill, knowledge, and experience to warrant a permit.

Despite the growth of opportunities for formal training, many North American banders gain their initial skills, knowledge, and experience through informal training, defined here as working in the presence of a trainer until the latter is satisfied with the consistency and correctness of data collection techniques and procedures. The purpose of this study was to investigate the outcomes of informal training, by comparing results to the minimum performance standards in use at the time of the study (1993), prior to development of NABC materials. Although these criteria are now largely outdated, the study demonstrates the importance of both training and evaluation procedures in ensuring accurate and consistent results.

METHODS

Research took place in 1993 at Beaverhill Bird Observatory in Alberta, Last Mountain Bird Observatory in Saskatchewan, and Delta Marsh Bird Observatory in Manitoba. Informal training varied among these stations, but in all cases the trainer did not allow the trainee to collect data alone until the trainee had achieved a high degree of agreement with the trainer. Quantifying length of the training period was often difficult, because checking diminished gradually in most cases. For each trainee, we obtained an estimate of total experience, and an estimate (from the trainer) of the time the trainee had access to the trainer. I chose to define training period as time spent in proximity to the trainer, because this could be most readily quantified.

Each trainer–trainee combination independently examined and collected data on the same birds. Number of birds measured by each trainer–trainee combination varied from 37 to 171. All data were collected in August and September, so participants were not usually able to use cloacal protuberance and brood patches as an indication of the sex or age of the birds. All stations used a five-point fat scale. Two stations used a three-class and one a six-class skull ossification scale. All participants recorded data without input from others (usually out of sight from one another). No discussion of birds being handled was allowed for the entire length of the experiment. For the purposes of this study, it was assumed that the trainer had correctly classified, assessed, and measured the bird.

For measurements, I calculated the average of the absolute deviations of trainee data from those of the trainer, and divided deviation by the average value achieved by the trainer. For categorical scores (fat and skull), I determined the proportions of cases in which a trainee scored the bird the same as the trainer (agreement), differed by one class, or differed by two classes. For age and sex I calculated the

proportion of cases in which a trainee scored the bird exactly the same as the trainer (agreement).

Scores were assessed by comparison to minimum performance standards suggested by C. J. Ralph (pers. comm.), which were developed in 1993 for a one-week training course to teach banding skills. These criteria are shown in Table 1. However, some errors are less important than others, and thus I also determined whether errors were "critical" or "non-critical." A trainee classifying a bird as "unknown age" (or sex) when the trainer felt able to classify to an age or sex category was a non-critical error, whereas errors when trainer and trainee assign opposing age or sex classes were critical. For skulling, a class error within the hatch year categories was considered non-critical.

RESULTS

Quantitative Measures

Analysis of quantitative measurement differences were limited to wing chord. One trainee was in the "top" category, using the standards in Table 1, and the rest were comfortably within the "pass" category (Table 2). The bulk of the birds measured were small passerines with wing chords less than 100 mm, so any error was almost sure to put the trainee in the pass rather than top category. Most errors were similar in magnitude to the amount of variation typical of an individual repeatedly measuring the same bird. There was no relationship to the amount of time the trainee had spent in proximity to the trainer or to overall length of experience. Wing measurement appears to be a skill that is learned quickly, and the skill is retained well after contact with the trainer is over.

Qualitative Measures

Species

Correct identification of species ranged from 98 to 100%. Two of the four errors committed were transcription errors, with the trainee writing the name of the previous species instead of the species being processed. The other two errors involved confusion between Least (*Empidonax minimus*) and Alder (*E. alnorum*) flycatchers. Examination of measurements collected by the trainer and application of formulas showed that the trainees made the wrong decisions because they did not collect all the necessary data. Using the Table 1 standard of 100% to pass, there were two passing and three failing individuals.

Age

Using 100% as the pass score (Table 1), no trainee achieved a passing score for assigning age (Table 3). One trainer–trainee combination did agree on the age of 99% of the birds. The only disagreement was a bird classed as unknown age by the trainee. Of the remaining four banders, three achieved scores in excess of 80% and one failed by a wide margin. Most errors by these four banders were of a critical nature (an adult bird called hatch year or vice versa), rather than non-critical (an adult or hatching year called unknown age).

Sex

No trainee achieved a perfect score on assigning sex (Table 3), so all failed according to the standards in Table 1. One trainee achieved a score of 98%, three more achieved scores above 80%, and one failed by a wide margin. Most errors were of a non-critical nature, in which the trainee classed the bird as unknown sex while the trainer classified it as known sex. However, every trainee made at least one critical error.

Skull

No individual attained a top score for correct skull classification (Table 4) according to the standards in

Table 1. Minimum performance standards for banders expressed as acceptable percent errors or concurrence between trainee and trainer

	Measurement	Species, age, sex	Qualitative (skull, fat)	
	% error	% agree	% agree	% differ by one class
Top	< 1	100	> 95	< 5
Pass	> 1 to 3	100	80–95	< 20
Marginal/ fail	> 3 to < 5	n/a[a]	50–80	20–40
Definite fail	> 5	< 100	< 50	> 50

Notes: Standards are those suggested by C. J. Ralph (pers. comm.) in 1993. All rates of agreement or error are in reference to answer as determined by the trainer.

[a]Anything less than 100% agreement for these categories was considered a failure so there is no marginal score for these skills.

TABLE 2. TRAINEE SCORES FOR WING MEASUREMENT AND SPECIES IDENTIFICATION

Trainer/Trainee (N)	Wing measurement % deviation	Species identification % agreement[a]
A / B (171)[b]	1.77	**99**
B / C (169)	1.62	**98**
D / E (100)	0.72	**99**
F / G (86)	1.90	100
F / H (37)	1.40	100

[a] Bold marks are failures by standards in Table 1.
[b] Sample size for these two skills for each pairing of personnel appears in parentheses.

Table 1. One passed, two achieved marginal scores, and two failed. With the exception of bander H, most errors were of a non-critical nature (differed in class within bird of the year categories), and these birds would have been aged correctly on the basis of skull. Using critical and non-critical classification for errors produces somewhat different results than does "differences of one class." Judging on the basis of Table 1, the number of serious errors made by G and H would have been underestimated, and the number of serious errors made by B would have been overestimated (Table 4).

Observer H had a high number of errors in skulling (Table 4). According to F (the trainer), H appeared to be skulling well at the end of the training period but had not subsequently asked for confirmation on many birds when trainer and trainee were in proximity. There was some parallel between scores on age and skull for H. This did not hold true for B, C, and E who seemed to have acceptable skulling ability (at least 85% agreement or non-critical errors), but did not assess ossification on some birds, and this is where most of their aging errors occurred. Skulling every bird would probably have improved their age classification performance.

Fat

There was one pass and three marginal scores (Table 5). Almost all errors, even by the failing individual H, were within a class of the trainer's determination.

Training levels

All the trainees had achieved a high degree of agreement with their trainers after initial training

TABLE 3. TRAINEE SCORES FOR AGE AND SEX DETERMINATION

Trainer/	Age			Sex		
Trainee (N)	% agree	% non-critical error	% critical error	% agree	% non-critical error	% critical error
A / B (171)[a]	**84**[b]	1	15	**70**	29	1
B / C (169)	**92**	0	8	**98**	0	2
D / E (100)	**99**	1	0	**85**	11	4
F / G (86)	**93**	1	6	**86**	9	5
F / H (37)	**65**	0	35	**84**	8	8

Notes: Values are expressed as percent of agreement, non-critical, and critical error.
[a] Sample size for these two skills for each pairing of personnel appears in parentheses.
[b] Bold marks are failures by standards in Table 1.

TABLE 4. TRAINEE SCORES FOR SKULL OSSIFICATION

Trainer/Trainee (N)	% agree	% non-critical error	% critical error	% differ by one class	% differ by two classes
A / B (52)[a]	**48**[b]	40	12	37	15
B / C (51)	63	33	4	33	4
D / E (87)	74	26	0	26	0
F / G (76)	80	13	7	19	1
F / H (37)	**46**	24	30	49	5

Notes: Values are expressed as percent agreement and percent by error type.
[a] Sample size for this skill for each pairing of personnel appears in parentheses.
[b] Bold marks are failures by standards in Table 1.

TABLE 5. TRAINEE SCORES FOR FAT ASSESSMENT

Trainer/Trainee (N)	% Agree	% Differ by one class	% Differ by two classes
A/B (171)[a]	79	19	2
B/C (169)	87	12	1
D/E (100)	61	31	8
F/G (86)	56	41	3
F/H ((37)	**40**[b]	49	11

Notes: Values are expressed as percent agreement and percent by error type.
[a] Sample size for this skill for each pairing of personnel appears in parentheses.
[b] Bold marks are failures by standards in Table 1.

(prior to this experiment). Experience gained subsequent to training (as measured by number of birds banded), and the period of long term access to the trainer following training, differed among those tested (Table 6). There was only one trainee (B) who was given a defined period of training and then banded alone thereafter.

It was recognized from the onset that it would be difficult to separate the influences of training and experience, because both are often acquired together and quantifying them in a meaningful way is difficult. The small sample size precludes quantitative analysis. Although data for the first three individuals in Table 6 suggest that access to a trainer beyond the first intensive period may be a factor in long term performance, this was not consistent. For example, bander H had a very long period of access to a trainer, but the worst score.

Results from bander B suggest that practice alone does not increase performance (Table 6). DeSante et al. (*this volume*) also presented data indicating that experience of banders does not necessarily ensure a higher degree of accuracy.

Discussions with F, the trainer of G and H, revealed that personality or temperament may be an important factor in training effectiveness. For example, trainee G was trained for a short time but was extremely cautious. Trainee G frequently asked questions of the trainer and spent a lot of time reading source and reference materials. Trainee H did well in initial training and testing, but rarely asked questions during the extensive period following training when the trainer F was accessible but not actively probing and testing H.

TABLE 6. ACCESS TO TRAINER, EXPERIENCE AND PERFORMANCE OF BANDING TRAINEES

Bander	Access to trainer (days)	Experience (birds banded)	Cumulative score (out of 600)
B	10	>3,000	478
C	65	~1,000	536
E	55	~2,000	518
G	29	~2,000	513
H	60	~2,000	433

DISCUSSION

No individual attained a fully satisfactory performance level based on the standards in Table 1. Several individuals had mainly errors of a non-critical nature, which was not reflected in the Table 1 scoring system that was in use at the time of the study. Current standards for performance assessment are quite different. The NABC does not treat all errors as equal, and although the council sets a high standard, it does not expect performance of 100% in aging and sexing birds. NABC standards also penalize critical errors more harshly than non-critical errors, because classing a bird as unknown age or sex is preferable to categorizing it incorrectly. Determination of age and sex is often based on subtle plumage characteristics, and it is to be expected that trainees will record a greater number of unknowns than trainers. Indeed, a trainee who rarely uses the "unknown" category may be overconfident, and probably should be rechecked for errors (M. McNicholl, pers. comm.). Nonetheless, the NABC does impose some penalty for non-critical errors made during testing, to encourage precision when a true determination is possible.

Despite the improvement of training guides and development of performance standards for certification, results in this paper indicate that individual differences among banders can readily arise and be promulgated. A good example of this is the case of banders B and C in this study. Bander B was given a short period of intensive training and then banded for a summer. The next year, B trained C, and the two worked together for the summer. It appears that because C had constant access to B prior to testing in this study, there was a high degree of agreement with B during the test. In fact, C was the only individual who came close to achieving a passing score. It appears B

had done a very good job of passing on information to C, which was the testing criterion in this study. However, B's score indicates that the information passed on to C was incomplete or incorrect.

Results of this study indicate that trainers should achieve a common standard before we rely on agreement of trainer and trainee results as the test of competency for new banders. Without this initial standardization, we will be perpetuating high variability in standards, because trainees reflect their trainer's skills. Moreover, it is important to recognize that learning and evaluation must not cease at the end of the training period. Recommendations to address these issues include the following:

1. More banding stations should undertake evaluations of their training effectiveness. This may clarify which factors most influence performance, and identify weaknesses in training programs. Especially needed is development of a schedule for follow-up spot checking after initial training has been completed.

2. Trainers should attend regional or national workshops so that all trainers teach from a similar standard. Contact and verification among trainers in a region should take place at least annually to maintain consistency.

3. Station personnel (regardless of experience) should periodically compare results, and immediately discuss sources of variation to iron out any problems revealed. For example, as a result of this study, F gave H a refresher course and they began regular comparisons, which showed a much higher level of agreement.

4. The role of trainer and trainee temperament should be given consideration in designing and carrying out training and assessment.

ACKNOWLEDGMENTS

The cooperation of the Beaverhill Bird Observatory, Last Mountain Bird Observatory, and Delta Marsh Bird Observatory is greatly appreciated. They are to be commended for permitting this experiment and for their positive attitude in using the results to improve their programs. I especially thank the anonymous trainers and trainees. R. D. Dickson, E. H. Dunn, E. Hayakawa, M. K. McNicholl, and C. J. Ralph offered helpful comments on the manuscript.