

Repeatability of subjective grading in fur animals

III. Grading of live blue foxes in different environmental conditions

HILKKA KENTTÄMIES and KERSTIN SMEDS

KENTTÄMIES, H. & SMEDS, K. 1992. **Repeatability of subjective grading in fur animals. III. Grading of live blue foxes in different environmental conditions.** *Agric. Sci. Finl.* 1: 315-322. (Univ. Helsinki, Dept. Anim. Breed., SF-00710 Helsinki, Finland.)

Association between repeatedly scored body size and fur characteristics were studied in live blue foxes. Gratings in cages and outside cages in lamplight and daylight were also compared. Colour tended to be easier and clarity more difficult to evaluate than the other traits. Differences between judges in accuracy of grading were greater than between various grading environments. The grading was more reliable outside cages than within cages. The most uniform results were obtained when the same judge graded the animals in the same environmental conditions.

Key words: fur characteristics, repeated evaluation, cage grading, corridor grading

Introduction

Reliability of subjective grading among fur bearing animals is highly dependent on the animal species, the skill of the judge and the environmental circumstances (JONSSON 1971, REITEN 1977 b, JEŻEWSKA and MACIEJOWSKI 1982, KENTTÄMIES and KÄYHKÖ 1992, KENTTÄMIES and SMEDS 1992). In practice, numerous farmers perform the grading of live animals themselves, at least in grouping the animals for breeding or pelting. The grading inside cages is commonly used as the first rough evaluation in selecting animals for pelting. A more detailed grading of the animals for breeding is often carried out outside cages by an adviser.

In previous studies with live minks and foxes, fairly low heritability estimates have been found for some subjectively scored traits, such as the quality traits (JONSSON 1971, REITEN 1977 a, KENTTÄMIES 1988, LAGERKVIST and LUNDEHEIM

1990) and clarity (ROSBERG and OLAUSSON 1978, KENTTÄMIES 1988). This may be due to the fact that the additive genetic proportion of the total variation for the traits is low or that the data used for the estimation of heritabilities have been less suitable for these purposes. One reason may be that the traits have been uncertainly determined or they have been difficult to judge reliably. The traits are, however, used for the evaluation and selection of breeding animals.

This paper is one in a series concerning the accuracy of grading in fur bearing animals. The reliability of grading for live foxes was discussed in Part I (KENTTÄMIES and KÄYHKÖ 1992), and that for mink and blue fox pelts in Part II (KENTTÄMIES and SMEDS 1992). The purpose of the present study was to ascertain the reliability of grading for repeatedly scored traits in live blue foxes. The aim was also to compare evaluations in different grading conditions conducted by different judges.

Material and methods

An experiment for repeated grading in different environmental conditions, was carried out at the Agricultural Research Centre, the Fur Farming Research Station.

In November 1989, a total of 77 live blue foxes (35 males and 42 females) were independently judged by three trained persons, two advisers from the Finnish Fur Animal Breeders' Association and a research technician from the Research Station. Each judge evaluated the blue foxes three times in succession in cages in daylight and four times in a corridor where the judging was done two times in daylight and two times in lamplight. The last grading, outside cages in lamplight, was postponed to the third grading day. The procedure for evaluation is presented as follows:

Date of judging	Grading condition	Times of judging
22.11.	In cages, daylight	3
23.11.	Outside cages, daylight	2
23.11.	Outside cages, lamplight	1
24.11.	Outside cages, lamplight	1

The animals (with a few exceptions) were graded a total of 21 times. The following five traits were evaluated in each grading: body size, colour, density of underfur, density of guard hairs, and clarity. A scale from 1 to 5 was used where 1 denotes smallest, darkest or poorest, and 5 largest, lightest or best.

Weather

The temperature, humidity and lightness in the shelter house were measured at least three times a day. The temperature varied from -1 to -5°C, while the humidity remained at 100 % during the experiment. The quality of light varied in the course of the trial. For the first day the sun was shining, the second and third day were cloudy and occasionally snow or wet snow was falling. On the first day at the start of grading, the intensity of the daylight was 170 lux. At noon a value of 5000 lux was measured on the sunny

side and 650 lux on the shaded side of the house, in the afternoon (at 2 pm) 600 lux was found on the sunny side and 150 lux on the shaded side of the house. On the second and third day, the intensity of daylight was 50 lux in the morning, 130 lux at noon and 20 lux in the afternoon while an intensity of 240-320 lux was measured when a lamp was used.

Statistical methods

The material was edited and analysed using a statistical program WSYS (VILVA 1989). Factors affecting the five traits were studied using the analysis of variance with the following model:

Model 1:

$$Y_{ijklm} = \mu + a_i + b_j + c_k + d_l + e_{ijklm}$$

where Y_{ijklm} = individual observation

μ = general mean

a_i = effect of the i th judge; $i=1-3$

b_j = effect of the j th grading condition; $j=1-3$ (1=in cages in daylight, 2=outside cages in daylight, 3=outside cages in lamplight)

c_k = effect of the k th time of judging; $k=1-3$ (in cages), $k=1,2$ (outside cages both in daylight and lamplight)

d_l = effect of the l th sex of animal; $l=1,2$ (1=females, 2=males)

e_{ijklm} = random error

The effects of the judge, grading condition, time of grading and sex of animal were considered to be fixed. Interactions between these factors were also studied. Data were also analysed separately by judges and grading conditions. The effect of light intensity and the interactions between judge and the intensity of light on the traits were studied within various grading conditions. The data were also analysed separately for males and females.

Accuracy of grading was defined as a repeatability coefficient for the repeated scorings. The repeatabilities were analyzed using random and mixed models. In the random model the effects of judge and grading environment were ignored, and the data were analyzed using the following model:

Model 2:

$$Y_{no} = \mu + f_n + e_{no}$$

where f_n = random effect of the n th animal with all other factors as described in Model 1.

The foregoing model was also used in analyzing repeatabilities by judges and grading conditions.

Repeatabilities within judges and grading conditions were analysed according to the following model:

Model 3:

$$Y_{ijno} = \mu + a_i + b_{ij} + f_{ijn} + e_{ijno}$$

where b_{ij} = random effect of the j th condition within i th judge

f_{ijn} = random effect of the n th animal within i th judge and j th condition with all other factors as described in Model 1.

Repeatabilities were analysed separately within grading conditions according to the following model:

Model 4a:

$$Y_{ino} = \mu + a_i + f_{in} + e_{ino}$$

where f_{in} = random effect of the n th animal within i th judge with all other effects as described in Model 1.

In order to analyse repeatabilities separately within judges, another model, Model 4b was applied. The fixed effects of the judge were substituted by grading condition.

Results and discussion

Means and standard deviations

The means of the traits (Table 1) were near the median of the scale, 3, while the extreme scores (1 and 5) were seldom used. Clarity was evaluated

Table 1. Number of observations (N), means and standard deviations (SD) for scores of body size and the fur traits in live blue foxes.

Trait	N	Mean	SD
Body size, females	873	2.78	0.76
Body size, males	730	3.39	0.87
Colour	1604	3.02	0.72
Underfur density	1604	2.89	0.83
Guard hair density	1604	3.23	0.82
Clarity	1604	3.40	0.71

according to a higher and more narrow scale than the other fur traits. There existed less variation in clarity and colour (SD=0.71–0.72) as compared with the other traits (SD=0.76–0.87). Large variations in the traits may be due to the fact that some of the animals were on a restricted feeding regime which resulted in changes in appearance.

Factors affecting the traits

The judge was found to have the greatest effect on the grading of the traits. Differences between judges appeared to be highly significant for each trait. Significant differences between grading environments were found for the traits, apart from body size (Table 2). The colour of the same animals tended to look lighter in lamplight than in daylight. The cleanest colours were found in the corridor in daylight. In the cage grading the animals were given higher scores for density of underfur as compared with the grading in the corridor. A reverse situation was found for the density of guard hairs.

There were statistically significant interactions between judge and grading condition in colour, density of underfur and density of guard hairs and some judges could not totally follow the same scale in each environment.

The scores given on different times of grading appeared to be much the same. Nevertheless, when the latter evaluation outside cages in lamplight was postponed to the following day, differences between the two assessments for body size as well as interactions between judge and time of grading for colour, underfur density and guard hair density tended to be significant.

In cloudy weather, slight differences observed in daylight intensity did not seem to affect the traits. In sunny weather, great differences which appeared in the intensity of daylight were, however, found to significantly affect the scores for colour, underfur density and guard hair density. Besides these traits, significant interactions between judge and daylight intensity appeared for body size. Nevertheless daylight intensity was not found to affect the grading of clarity.

Table 2. Effects of judge and grading condition on scores for body size and the fur traits in live blue foxes.

Factor	N	Least Squares Means				
		Body size	Colour	Underfur density	Guard hair density	Clarity
Judge						
1	528	3.12	3.02	3.07	3.02	3.72
2	537	2.67	3.41	2.44	2.92	3.23
3	538	3.46	2.72	3.21	3.83	3.29
		***	***	***	***	***
Condition						
1	691	3.07	2.92	2.99	3.15	3.33
2	450	3.10	3.02	2.83	3.30	3.50
3	462	3.08	3.20	2.90	3.31	3.41
		NS	***	**	***	***

Condition: 1 = in cages in daylight, 2 = outside cages in daylight, 3 = outside cages in lamplight

*** $P < 0.001$, ** $P < 0.01$, NS = not significant

In all the data as well as in the analyses done by judges, males tended to obtain significantly higher points for body size and density of underfur as compared to females.

Repeatability of grading

Average associations between the repeated evaluations are given in Table 3. Higher repeatabilities were obtained after eliminating the differences and possible interactions between environmental conditions and judges (Within judge, condition and animal) as compared with those obtained from the animal components only (Within animal). Due to the obvious effects of judge and environment on the traits, the previously mentioned repeatability coefficients are considered more reliable than the uncorrected ones.

Colour was found easier to judge than the other traits. On the other hand, clarity and density of guard hairs seemed to be considerable difficult to grade. These results are in accordance with previous reports for minks (JONSSON 1971, REITEN 1977 b, JEZEWSKA and MACIEJOWSKI 1982, KENTTÄMIES and SMEDS 1992), for blue foxes (JEZEWSKA and MACIEJOWSKI 1982, KENTTÄMIES and SMEDS 1992),

and for silver foxes (JEZEWSKA and MACIEJOWSKI 1982, KENTTÄMIES and KÄYHKÖ 1992).

Differences between environments

The highest repeatabilities were mostly obtained when the grading was conducted in the corridor in daylight, and the lowest ones in cages (Table 4). Similar repeatabilities were, however, obtained in cages and in the corridor in lamplight for colour and underfur density, the most reliable repeatabilities being in the corridor in daylight ($r = 0.63$, 0.64 vs. 0.76 for colour; $r = 0.51$, 0.48 vs. 0.71 for

Table 3. The repeatability coefficients \pm standard errors for scores of body size and the fur traits in live blue foxes.

Trait	Repeatabilities \pm standard errors	
	Within animal	Within judge, condition and animal
Body size, females	0.39 \pm 0.06	0.58 \pm 0.03
Body size, males	0.34 \pm 0.06	0.59 \pm 0.03
Colour	0.48 \pm 0.04	0.67 \pm 0.02
Underfur density	0.35 \pm 0.04	0.56 \pm 0.02
Guard hair density	0.26 \pm 0.04	0.56 \pm 0.02
Clarity	0.16 \pm 0.03	0.53 \pm 0.03

underfur density). The cage grading produced clearly lower repeatabilities for clarity than the corridor grading both in lamplight and daylight ($r = 0.46$ vs. 0.60 and 0.59).

In the cage grading, a lower coefficient of repeatability for colour was found in the present study ($r = 0.63$) than in a previous one for silver foxes ($r = 0.74$) (KENTTÄMIES and KÄYHKÖ 1992). However, similar coefficients were found for body size ($r = 0.56$ for blue foxes vs. 0.55 for silver foxes).

The grading of blue foxes in the corridor in daylight, produced higher repeatabilities in the present study than in a previous study where two different groups of judges with varying skills performed the grading (JEŻEWSKA and MACIEJOWSKI 1982). In each study fairly high repeatabilities were obtained for colour ($r = 0.76$ vs. 0.61 and 0.67 in JEŻEWSKA and MACIEJOWSKI 1982). In the present study a high coefficient was obtained also for the density of underfur ($r = 0.71$), while lower and varying ones were found by JEŻEWSKA and MACIEJOWSKI 1982 ($r = 0.27$ and 0.51).

The use of a lamp was expected to increase the reliability of grading, higher repeatabilities were, however, obtained in daylight. Equal repeatabilities were found only in the grading for clarity. Fairly low accuracy received in lamplight may be partly due to the fact that the second judging was postponed to the following day. This was found to change the scale of grading which could at the same time negatively affect the correlations. In addition, the judges had not become accustomed to evaluate animals using a lamp.

The evaluations in cages and in the corridor in daylight were repeated on the same day. During the corridor grading in daylight, enough light was available due to a clear covering of snow, in spite of the cloudy weather. In these conditions trained judges succeeded in correcting for slight changes in lighting. No previous studies concerning the comparison of live animal grading in different environments seem to have been undertaken. However, JONSSON 1971 and REITEN 1977 b found obvious differences in correlations between repeated grad-

Table 4. The repeatability coefficients \pm standard errors for scores of body size and the fur traits in live blue foxes analysed by grading conditions.

Trait	Repeatabilities \pm standard errors		
	Cages	Corridor	
	daylight	daylight	lamplight
Body size, females	0.45 \pm 0.05	0.72 \pm 0.04	0.64 \pm 0.05
Body size, males	0.47 \pm 0.06	0.73 \pm 0.05	0.65 \pm 0.06
Colour	0.63 \pm 0.03	0.76 \pm 0.03	0.64 \pm 0.04
Underfur density	0.51 \pm 0.04	0.71 \pm 0.03	0.48 \pm 0.05
Guard hair density	0.45 \pm 0.04	0.70 \pm 0.03	0.56 \pm 0.04
Clarity	0.46 \pm 0.04	0.59 \pm 0.04	0.60 \pm 0.04

Table 5. The repeatability coefficients \pm standard errors for scores of body size and the fur traits in live blue foxes analysed by judges.

Trait	Repeatabilities \pm standard errors		
	Judge 1	Judge 2	Judge 3
Body size, females	0.69 \pm 0.04	0.48 \pm 0.06	0.42 \pm 0.06
Body size, males	0.60 \pm 0.06	0.60 \pm 0.06	0.56 \pm 0.06
Colour	0.68 \pm 0.03	0.68 \pm 0.03	0.65 \pm 0.03
Underfur density	0.53 \pm 0.04	0.64 \pm 0.04	0.53 \pm 0.04
Guard hair density	0.61 \pm 0.04	0.59 \pm 0.04	0.47 \pm 0.05
Clarity	0.67 \pm 0.03	0.30 \pm 0.05	0.20 \pm 0.05

ings for the corresponding traits among the same live and pelted minks.

Differences between judges

Differences between judges in the repeatability coefficients existed for the most traits (Table 5). The greatest differences between judges were found in clarity, the repeatabilities were from 0.20 to 0.67. A previous knowledge of the animals appeared to be of advantage, as the highest coefficients of repeatability for most of the traits were received by the judge working at the experimental farm.

The repeatability coefficients attained by judges in different environments varied from 0.17 to 0.83 (Table 6). The difficulty involved in subjective evaluation is characterized by the fact that only seldom did the same judge produce the highest coefficients in each environment. By far the most uniform repeatabilities were obtained in the corridor in daylight, while the largest differences between judges appeared when using a lamp.

Using the cage grading, similar differences between judges in their ability to grade for colour were found as previously determined among silver foxes (KENTTÄMIES and KÄYHKÖ 1992). The repeatabilities obtained in the present study tended, however, to be lower than those found for silver foxes ($r = 0.57$ to 0.72 for blue foxes vs. 0.68 to 0.82 for silver foxes). Smaller differences between judges were found for body size ($r = 0.48$ to 0.66 for blue foxes vs. 0.51 to 0.59 for silver foxes). Also in the grading outside cages, JEZEWSKA and MACIEJOWSKI (1982) found remarkable differences between different people in their competence of grading for total evaluation in blue foxes and minks, lower differences were found for silver foxes.

Lower repeatabilities for colour and clarity were obtained in the present study as compared with an earlier study for blue fox pelts graded by the same judges (KENTTÄMIES and SMEDS 1992). In live animals graded outside cages using a lamp, the coefficients of repeatability for colour varied from 0.52

Table 6. The repeatability coefficients \pm standard errors for scores of body size and the fur traits in live blue foxes analysed by judges in different environments.

Trait	Repeatabilities \pm standard errors								
	Cages, daylight			Corridor, daylight			Corridor, lamplight		
	Judge 1	Judge 2	Judge 3	Judge 1	Judge 2	Judge 3	Judge 1	Judge 2	Judge 3
Body size, females	0.57 \pm 0.08	0.30 \pm 0.10	0.24 \pm 0.10	0.75 \pm 0.07	0.71 \pm 0.08	0.61 \pm 0.10	0.82 \pm 0.05	0.46 \pm 0.12	0.46 \pm 0.12
Body size, males	0.47 \pm 0.10	0.38 \pm 0.11	0.50 \pm 0.10	0.74 \pm 0.08	0.77 \pm 0.07	0.69 \pm 0.09	0.70 \pm 0.09	0.78 \pm 0.07	0.53 \pm 0.12
Colour	0.57 \pm 0.06	0.72 \pm 0.05	0.60 \pm 0.06	0.82 \pm 0.04	0.76 \pm 0.05	0.72 \pm 0.05	0.77 \pm 0.05	0.52 \pm 0.08	0.64 \pm 0.07
Underfur density	0.52 \pm 0.06	0.60 \pm 0.06	0.44 \pm 0.07	0.64 \pm 0.07	0.73 \pm 0.05	0.75 \pm 0.05	0.49 \pm 0.09	0.55 \pm 0.08	0.41 \pm 0.10
Guard hair density	0.55 \pm 0.06	0.41 \pm 0.07	0.32 \pm 0.07	0.71 \pm 0.06	0.72 \pm 0.05	0.68 \pm 0.06	0.69 \pm 0.06	0.69 \pm 0.06	0.33 \pm 0.10
Clarity	0.61 \pm 0.06	0.27 \pm 0.07	0.17 \pm 0.07	0.69 \pm 0.06	0.37 \pm 0.10	0.34 \pm 0.10	0.75 \pm 0.05	0.32 \pm 0.10	0.21 \pm 0.11

to 0.64, and for clarity from 0.21 to 0.32. In pelts, the respective coefficients were 0.77 to 0.90 for colour and 0.54 to 0.66 for clarity. Therefore it can be stated that not even a trained judge is capable of evaluating live blue fox fur as reliably as their pelts.

Conclusions

A series of experiments investigating the accuracy of subjective grading was conducted on fur bearing animals. Repeatability of grading for live foxes was discussed in Part I (KENTTÄMIES and KÄYHKÖ 1992), for mink and blue fox pelts in Part II (KENTTÄMIES and SMEDS 1992) and for live blue foxes in different grading conditions in the present study (Part III).

In scoring living and pelted blue foxes, the same judges obtained higher repeatabilities for the respective traits in pelts as compared with live animals. Therefore it can be concluded that evaluating live fur animals is not as reliable as evaluating their pelts. It seems that live silver foxes were easier to judge than blue foxes. On the other hand blue fox pelts seemed to be easier to evaluate than minks. In addition, there was a tendency for more reliable grading among some colour types as com-

pared with some others.

In each paper differences were found between judges in the scales used for judging. In addition, differences between the repeatabilities obtained by various persons were obvious. The accuracy of grading was greatly affected by the grading environment, grading in the corridor being more accurate than in cages. However, the first assessment of appearance can be done in cages in order to group the animals for breeding or pelting.

In each paper colour was found to be a fairly easy trait to evaluate, while clarity tended to be one of the most difficult ones. Scoring for individual quality traits, ie. density of guard hairs, and particularly density of underfur seemed to be easier than for overall fur quality. Subjective evaluation for body size seemed to be fairly easy, although a higher accuracy was obtained by measuring the length of pelt. Grading for pelt defects appeared to be fairly difficult.

Acknowledgements.

Pekka Petman and Simo Paavola, Finnish Fur Breeders' Association and Pekka Toikkonen, the Research Fur Farm of the Agricultural Research Centre are acknowledged for carefully grading the animals. Financial support was received from the Technology Development Centre (TEKES).

References

- JEŻEWSKA, G. & MACIEJOWSKI, J. 1982. Repeatability of exterior trait estimations in different species of fur animals. 33rd EAAP, Leningrad. 9 p.
- JONSSON, M.B. 1971. Variasjonsårsaker for noen produksjonsegenskaper hos mink. Meld. Norg. Landbr.høgsk. 50, no. 6. 47 p.
- KENTTÄMIES, H. 1988. Heritability of body size and fur quality of foxes *Vulpes vulpes*. In: Murphy, B.D. & Hunter, D.B. (eds.). Biology, Pathology and Genetics of Fur Bearing Animals. (Proc. IV Int. Congr. Fur Anim. Breed. Toronto, Canada, p. 548-556.
- & KÄYHKÖ, H. 1992. Repeatability of subjective grading in fur animals. I. Grading of live silver foxes. Agric. Sci. Finl. 1: 303-307.
- & SMEDS, K. 1992. Repeatability of subjective grading in fur animals. II. Grading of mink and blue fox pelts. Agric. Sci. Finl. 1: 309-314.
- LAGERKVIST, G. & LUNDEHEIM, L. 1990. Fur quality traits in Standard mink-Price relationships, heritabilities and genetic and phenotypic correlations. Acta Agric. Scand. 40: 367-376.
- ROSBERG, S.O. & OLAUSSON, A. 1978. Nedärvning av vikt och päls-egenskaper hos standard och safirmink. Våra Pälsdjur 49: 173-177.
- REITEN, J. 1977 a. Arvbarhetsestimater for størrelse og pelsegenskaper hos mørk mink. Meld. Norg. Landbr.høgsk. 56, no. 16, 12 p.
- 1977 b. Korrelasjoner mellom gjentatte målinger og vurderinger av pels-egenskaper hos mørk mink. Meld. Norg. Landbr.høgsk. 56, no. 14, 15 p.
- VILVA, V. 1989. Statistical microcomputer program WSYS. Univ. Helsinki, Dept. Anim. Breed.

Manuscript received February 1992

Hilkka Kenttämies
Kerstin Smeds
University of Helsinki
Department of Animal Breeding
SF-00710 Helsinki, Finland

Kerstin Smeds
Present address:
Finnish Fur Breeders' Association
Box 5, SF-06101 Vantaa, Finland

SELOSTUS

Turkiseläinten subjektiivisen arvostelun toistettavuus III. Elävien sinikettujen arvostelu erilaisissa ympäristöoloissa

HILKKA KENTTÄMIES ja KERSTIN SMEDS
Helsingin yliopisto

Sinikettujen subjektiivisen arvostelun toistettavuutta tutkittiin Maatalouden tutkimuskeskuksen Turkiseläintuotannon tutkimusasemalla Kannuksessa. Kokeessa verrattiin arvostelun luotettavuutta, kun eläimet olivat arvostelun ajan häkissä tai ne nostettiin käytävälle. Häkissä eläimet arvosteltiin päivänvalossa, käytävällä ne arvosteltiin sekä päivänvalossa että lampunvalossa. Varjotalon lämpötila, kosteus ja valon voimakkuus mitattiin. Kolme arvosteluun tottunutta henkilöä arvosteli 77 sinikettua itsenäisesti 2-3 kertaa mainituissa olosuhteissa. Eläimiltä arvosteltiin koko, väri, massakkuus,

peittävyys ja värin puhtaus. Värin tummuuden arvostelu oli helpompaa ja värin puhtauden vaikeampaa kuin muiden ominaisuuksien arvostelu. Arvostelijoiden väliset erot olivat suurempia kuin eri olosuhteissa tapahtuvan arvostelun väliset erot. Arvostelu käytävällä oli luotettavampaa kuin häkissä. Suuret valon voimakkuuden vaihtelut häiritsivät värin ja laatuominaisuuksien arvostelua. Yhdenmukaisimmat tulokset saavutettiin, kun sama henkilö arvosteli eläimet samoissa olosuhteissa.