JDS Communications®
TBC; TBC

https://doi.org/10.3168/jdsc.2025-0930
Short Communication
Genetics

Cattle grow taller: Implications of outdated ordinal scores for genetic evaluations and selection?

Xiao-Lin Wu,[1,2]* Allicia Horn,[3] Ezequiel Nicolazzi,[1] and John B. Cole[1,4,5,6]

Abstract: In dairy cattle, conformation and linear-type traits are routinely recorded as categorical or ordinal phenotypes. Some traits originating from continuous biological measurements are discretized into categorical scores. These traits are moderately heritable and play a crucial role in breeding programs, serving as indicators of longevity, fertility, and overall lifetime productivity. However, over time, the underlying phenotypic distributions can evolve due to selection, management, and environmental changes. As a result, the mapping between biological measurements and categorical scores may become obsolete. This paper investigates phenotypic distribution shifts within ordinal scoring systems, using stature in Brown Swiss cattle as a case study. We present an analytical characterization of these shifts and their genetic consequences. A simulation study based on a real pedigree was conducted to quantify the extent of these effects and to explore methods for predicting these changes. The results emphasize periodic recalibration of threshold boundaries to keep the ordinal scoring systems aligned with current population distributions and safeguard the validity of genetic improvement programs. We anticipate that our findings will provide practical guidance for detecting, interpreting, and managing distributional shifts in categorical phenotypes within structured genetic evaluation and selection programs.

In dairy breeding and management, ordinal (linear-type) scoring provides a practically convenient and standardized framework for describing morphological and structural traits across herds and environments, typically on a 1 to 9 scale, with some systems extending to 50 or even 80 categories for finer discrimination. These traits are generally moderately to highly heritable (Veerkamp and Brotherstone, 1997; Manafiazar et al., 2016), serving as indirect indicators of important functional traits such as longevity, fertility, and overall lifetime productivity (VanRaden and Wiggans, 1995).

Linear scores are assigned according to well-defined biological attributes or biometric features. However, some traits originate from continuously measured variables that are subsequently transformed into categorical scores. For example, stature is inherently continuous but is routinely converted into a 1–9 scale, with each unit representing a 1-inch (**in**) interval. In the US Brown Swiss Association (**BSA**; Beloit, WI), for instance, the intermediate score of 5 corresponds to a stature of 56 in (142.24 cm). Animals shorter than 56 in receive scores from 1 (52 in) to 4 (55 in), whereas taller animals are assigned scores from 6 (57 in) to 9 (60 in).

Over time, however, the phenotypic distributions of such traits can shift. For example, modern cattle have become taller (Visscher and Goddard, 2011), largely due to correlated selection responses. Because height was rarely a primary breeding target, long-term selection for milk yield, angularity, and overall type possibly led to consistent upward drift in frame size. Height (stature) is correlated with udder depth, which is in turn associated with improved udder health. Taller cows carry their udders higher off the ground, reducing the risk of intramammary infections from environmental pathogens. In a recent survey of 5,279 Brown Swiss cows in 2024, we observed a mean ± SD stature score of 6.65 ± 1.45, indicating a clear upward shift in the underlying stature measurements, with the central tendency moving from the intended midpoint of 5 to approximately 7 on the ordinal scale. To address this discrepancy, the BSA announced a recalibration of the scoring system, raising all stature thresholds by 2 in, effective in the April 2025 genetic evaluation. This realignment of stature scores reflects current biological reality; however, the broader implications of such shifts in phenotypic distributions have not yet been systematically evaluated.

When categorical scaling systems are incorporated into genetic evaluations, key questions arise: How do phenotypic shifts in ordinal scores affect EBVs and heritability estimates? Furthermore, what are the genetic and selection consequences of re-anchoring the scoring scale? Therefore, the objectives of this study were to investigate phenotypic distribution shifts in ordinal scoring systems, using stature in US Brown Swiss (**BS**) cattle as a case study. We present an analytical framework to characterize these shifts and evaluate their consequences on phenotypic distributions, genetic evaluations, and selection outcomes.

First, how do ordinal scores change under a shifting phenotypic distribution? Let $\mathbf{y}$ denote a vector of continuous trait values that are later discretized into ordinal scores using the equal-interval (**EI**) binning method, the standard approach adopted by the BSA. Briefly, the EI approach divides the range of continuous phenotypic values into k ordered categories (bins) of equal width, with the first and last bins defined as open-ended intervals to accommodate extreme observations. The thresholds are defined as

[1]Council on Dairy Cattle Breeding, Bowie, MD 20716, [2]Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI 53706, [3]Brown Swiss Association, Beloit, WI 53511, [4]Department of Animal Sciences, Donald Henry Barron Reproductive and Perinatal Biology Research Program, and the Genetics Institute, University of Florida, Gainesville, FL 32611, [5]Department of Animal Science, North Carolina State University, Raleigh, NC 27607, [6]Department of Animal Biosciences, University of Guelph, Ontario, Canada N1G 2W1. *Corresponding author: nick.wu@uscdcb.com. Received November 03, 2025. Accepted January 03, 2026.

The list of standard abbreviations for JDSC is available at adsa.org/jdsc-abbreviations-26. Nonstandard abbreviations are available in the Notes.

$$\tau_0 = -\infty < \tau_1 < \ldots < \tau_{k-1} < \tau_k = +\infty.$$

For individual i, an ordinal score (s_i) is assigned as follows:

$$s_i = j \text{ if } \tau_{j-1} \le y_i < \tau_j, \text{ for } j = 1, \ldots, k.$$

Here, τ_{j-1} and τ_j define the lower and upper boundaries for the jth bin.

$$\text{Assume } y_i \sim \mathrm{N}\left(\mu, \sigma_e^2\right),$$

where μ is the population mean and σ_e^2 is the variance. The unconditional probability (**Pr**) of an observation falling into bin j is

$$\begin{aligned} p_j &:= \Pr\left(s_i = j\right) = \Pr\left(\tau_{j-1} \le y_i < \tau_j\right) \\ &= \Phi\left(\frac{\tau_j - \mu}{\sigma_e}\right) - \Phi\left(\frac{\tau_{j-1} - \mu}{\sigma_e}\right), \end{aligned} \quad [1]$$

where $\Phi\left(\cdot\right)$ denotes the cumulative distribution function of the standard normal distribution. The mean and variance of the ordinal score are then given by

$$\mathrm{E}\left(s\right) = \sum_{j=1}^{k} j p_j; \quad [2]$$

$$\mathrm{Var}\left(s\right) = \sum_{j=1}^{k} j^2 p_j - \left[\mathrm{E}\left(s\right)\right]^2. \quad [3]$$

For instance, let the current population mean be $\mu = 58$ in (147.32 cm) with SD of $\sigma_y = 1.87$ in (4.75 cm). The 9 categories are defined using a fixed 1-in interval for all bins except the first and the last categories: 1 ($y < 54.5$ in), 2 ($54.5 \le y < 55.5$ in), 3 ($55.5 \le y < 56.5$ in), 4 ($56.5 \le y < 57.5$ in), 5 ($57.5 \le y < 58.5$ in), 6 ($58.5 \le y < 59.5$ in), 7 ($59.5 \le y < 60.5$ in), 8 ($60.5 \le y < 61.5$ in), and 9 ($y \ge 61.5$ in). Here, the US customary units are retained for consistency with industry reporting. Under $y \sim N(58, 1.87^2)$, the proportions of observations falling in each group are approximately 3.06% (groups 1 and 9), 5.95% (groups 2 or 8), 12.2% (groups 3 or 7), 18.2% (groups 4 or 6), and 21.2% (group 5). These proportions are symmetric around the central category (group 5) and sum to 100% (see the top-left panel of the Graphical Abstract).

Now, suppose the current scores resulted from shifting upward from the outdated scores (denoted by s_i^*) by $\Delta = 2$ in (5.08 cm). Equivalently, this corresponds to shifting all interior thresholds downward by $\Delta = 2$ in (5.08 cm) while keeping the current continuous phenotypic values fixed. The probability of an observation falling into the jth category under the outdated scheme is

$$p_j^* := \Pr\left(s_j^* = j\right) = \Phi\left(\frac{\tau_j - \mu - \Delta}{\sigma_e}\right) - \Phi\left(\frac{\tau_{j-1} - \mu - \Delta}{\sigma_e}\right). \quad [4]$$

Applying these outdated thresholds yields the following proportions of observations across the 9 categories: 0.16% (group 1), 0.64% (group 2), 2.26% (group 3), 5.95% (group 4), 12.2% (group 5), 18.2% (group 6), 21.2% (group 7), 18.2% (group 8), and 21.2% (group 9). The downward-shifted thresholds thus lead to a disproportionate accumulation of observations in higher score groups (7–9), whereas lower groups (1–3) become sparsely populated (see the top-middle panel of the Graphical Abstract).

When the thresholds are further downscaled by $\Delta = 4$ in (10.16 cm), the proportions of observations across the 9 categories become 0.003%, 0.023%, 0.134%, 0.64%, 2.26%, 5.95%, 12.2%, 18.2%, and 60.6%, respectively. In this case, nearly two-thirds of the population is clustered in the highest category, while the lower categories become almost empty (see the top-right panel of the Graphical Abstract).

Failing to recalibrate the scoring system periodically leads to severe misalignment between ordinal scores and the underlying continuous phenotypes. The squared (and rank) correlations between the ordinal scores and the true continuous phenotypes decrease from 0.969–0.970 (0.987) for $\Delta = 0$ (in) to 0.924–0.926 (0.983) for $\Delta = 2$ (in), and then to 0.697–0.700 (0.879) for $\Delta = 4$ (in). The sharp decline in both metrics demonstrates a substantial loss of information and deterioration in phenotypic ranking accuracy as scoring thresholds become outdated.

Secondly, how do phenotypic shifts influence genetic evaluations? Consider the following linear animal model for the latent continuous phenotype (***y***):

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{u} + \boldsymbol{e}. \quad [5]$$

Here, **1** represents a vector of ones and $\boldsymbol{u} \sim N\left(0, \mathbf{A}\sigma_u^2\right)$ is a vector of additive genetic effects, where **A** is the numerator additive genetic relationship matrix and σ_u^2 is the additive genetic variance. The residual vector $\boldsymbol{e} \sim N\left(0, \mathbf{I}\sigma_e^2\right)$, with σ_e^2 denoting the residual variance and **I** the identity matrix. The vectors ***u*** and ***e*** are assumed to be mutually independent.

When fitting the same linear animal model to the ordinal scores ***s***, we write the following:

$$\boldsymbol{s} = 1\mu_s + \boldsymbol{u}_s + \boldsymbol{e}_s, \quad [6]$$

where $\boldsymbol{u}_s \sim N\left(0, \mathbf{A}\sigma_{u_s}^2\right)$ and $\boldsymbol{e}_s \sim N\left(0, \mathbf{I}\sigma_{e_s}^2\right)$, with $\boldsymbol{u}_s$ and $\boldsymbol{e}_s$ mutually independent.

The link between the latent-scale u and the score-scale s can be quantified analytically through the following best linear projection (**BLP**):

$$s_i = \alpha + \beta u_i + r_i, \quad [7]$$

where α and β are the intercept and regression coefficient, respectively, and r_i is a residual term with $\mathrm{E}(r_i) = 0$ and $\mathrm{Cov}(r_i u_i) = 0$. The BLP slope is defined by

$$\beta = \frac{\mathrm{Cov}\left(s_i, u_i\right)}{\mathrm{Var}\left(u_i\right)}. \quad [8]$$

Define the conditional mean function:

$$g(u_i) := \mathrm{E}(s_i \mid u_i) = \sum_{j=1}^{k} j \Pr(s_i = j \mid u_i)$$
$$= k - \sum_{j=1}^{k-1} \Phi\left(\frac{\tau_j - \mu - u_i}{\sigma_e}\right).$$

Because s_i depends on u_i only through $g(u_i) = \mathrm{E}(s_i \mid u_i)$, we have

$$\mathrm{Cov}(s_i, u_i) = \mathrm{Cov}[g(u_i), u_i].$$

Stein's lemma implies (Stein, 1956; Casella and Berger, 2002) that, valid for $u_i \sim N(0, \sigma_u^2)$,

$$\mathrm{Cov}[g(u_i), u_i] = \sigma_u^2 \mathrm{E}[g'(u_i)],$$

and therefore

$$\beta = \frac{\sigma_u^2 \mathrm{E}[g'(u_i)]}{\sigma_u^2} = \mathrm{E}[g'(u_i)].$$

Differentiating $g(u_i)$ with respect to u_i gives

$$g'(u_i) = \frac{1}{\sigma_e} \sum_{j=1}^{k-1} \phi\left(\frac{\tau_j - \mu - u_i}{\sigma_e}\right),$$

where $\phi(\cdot)$ is the standard normal probability density function (**PDF**).

Taking expectation over $u_i \sim N(0, \sigma_u^2)$ yields the Gaussian convolution:

$$\beta = \mathrm{E}[g'(u_i)] = \frac{1}{\sigma_y} \sum_{j=1}^{k-1} \phi\left(\frac{\tau_j - \mu}{\sigma_y}\right), \ \sigma_y^2 = \sigma_e^2 + \sigma_u^2. \quad [9]$$

Intuitively, when averaging a Gaussian PDF evaluated at linearly shifted arguments over a Gaussian random shift u, averaging "smooths" the PDF by convolving the 2 normals, which simply adds their variances: $\sigma_y^2 = \sigma_e^2 + \sigma_u^2$, and the prefactor $1/\sigma_y$ is the new normalizing scale.

The BLP [7] induces a natural projection-based variance decomposition on the score scale:

$$\mathrm{Var}(s_i) = \mathrm{Var}[\mathrm{E}(s_i \mid u_i)] + \mathrm{E}[\mathrm{Var}(s_i \mid u_i)] \approx \beta^2 \sigma_u^2 + \sigma_r^2, \quad [10]$$

where $\sigma_r^2 = Var(r_i)$. Define the BLP-implied additive genetic variance and BLP-implied residual variance on the ordinal scale as

$$\sigma_{u_s}^2 := \mathrm{Var}[g(u_i)] \approx \mathrm{Var}(\alpha + \beta u_i) = \beta^2 \mathrm{Var}(u_i) = \beta^2 \sigma_u^2; \quad [11]$$

$$\sigma_{e_s}^2 := \mathrm{E}[\mathrm{Var}(s_i \mid u_i)] \approx \mathrm{Var}(s_i) - \beta^2 \sigma_u^2. \quad [12]$$

Hence, the projection-based heritability on the ordinal scale is

$$h_s^2 := \frac{\sigma_{u_s}^2}{\sigma_{u_s}^2 + \sigma_{e_s}^2} \approx \frac{\beta^2 \sigma_u^2}{\mathrm{Var}(s)}. \quad [13]$$

The above approximation mappings depend on $\mu, \sigma_u^2, \sigma_e^2$, and $\{\tau_j\}$, but not on the structure of **A**. Thus, they provide a direct analytical approach to predict how phenotypic shifts rescale the score-breeding value relationship through β. However, the EBV precisions (e.g., their prediction error variances) still depend on **A** and the amount of information. These approximations are good as long as the score has many categories (k is not too small) and thresholds are not too extreme.

While these approximations are used in the present study, the exact definitions are the following:

$$\mathrm{Var}[g(u_i)] = \mathrm{E}[g(u_i)^2] - \{\mathrm{E}[g(u_i)]\}^2,$$

where both expectations are one-dimensional Gaussian integrals, typically computed by Gauss–Hermite quadrature or Monte Carlo. Likewise,

$$\mathrm{Var}(s_i \# u_i = u) = \mathrm{E}[s_i^2 \# u_i = u] - g(u)^2,$$

where $\mathrm{E}[s_i^2 \# u_i = u] = \sum_{j=1}^{k} j^2 \Pr(s_i = j \# u_i = u)$ and $\Pr(s_i = j \# u_i = u) = \Phi\left(\frac{\tau_j - \mu - u}{\sigma_e}\right) - \Phi\left(\frac{\tau_{j-1} - \mu - u}{\sigma_e}\right)$.

Now, consider downscaling the threshold by $\Delta > 0$. Employing the same BLP of s_i^* on u_i, the slope becomes

$$\beta^* = \frac{1}{\sigma_y} \sum_{j=1}^{k-1} \phi\left(\frac{\tau_j - \mu - \Delta}{\sigma_y}\right). \quad [14]$$

Similar to the expressions (11)–(13), the projected variances and heritability under the downscaled thresholds are

$$\sigma_{u_s}^{*2} \approx (\beta^*)^2 \sigma_u^2; \quad [15]$$

$$\sigma_{e_s}^{*2} \approx \mathrm{Var}(s^*) - (\beta^*)^2 \sigma_u^2; \quad [16]$$

$$h_s^{*2} \approx \frac{\left(\beta^*\right)^2 \sigma_u^2}{\mathrm{Var}\left(s^*\right)}. \quad [17]$$

The mean and variance of the ordinal scores under the downscaled thresholds are $\mathrm{E}\left(s^*\right) = \sum_{j=1}^{k} j p_j^*$ and $\mathrm{Var}\left(s^*\right) = \sum_{j=1}^{k} j^2 p_j^* - \left[\mathrm{E}\left(s^*\right)\right]^2$.

As Δ increases, probability mass accumulates in the uppermost category (ceiling effect), and both β^* and $\mathrm{Var}(s^*)$ decrease toward zero, leading to smaller score-scale heritability and reduced EBVs.

Rescaling the ordinal scores by a constant c changes variance components but not the heritability estimates. Let $\boldsymbol{l} = c\boldsymbol{s}$. The model becomes

$$\boldsymbol{l} = 1\mu_l + \boldsymbol{u}_l + \boldsymbol{e}_l, \quad [18]$$

where μ_l, $\boldsymbol{u}_l$, and $\boldsymbol{e}_l$ denote the overall mean, additive genetic, and residual effects on the rescaled ordinal scale. The rescaled variances are

$$\mathrm{Var}\left(\boldsymbol{l}\right) = c^2 \mathrm{Var}\left(\boldsymbol{s}\right); \quad [19]$$

$$\sigma_{u_l}^2 \approx c^2 \beta^2 \sigma_u^2 = c^2 \sigma_{u_s}^2; \quad [20]$$

$$\sigma_{e_l}^2 \approx c^2 \left[\mathrm{Var}\left(s\right) - \sigma_{u_s}^2\right] = c^2 \sigma_{e_s}^2. \quad [21]$$

However, heritabilities remain invariant under rescaling:

$$\mathrm{h}_l^2 = \frac{c^2 \sigma_{u_s}^2}{c^2 \sigma_{u_s}^2 + c^2 \sigma_{e_s}^2} = \frac{\sigma_{u_s}^2}{\sigma_{u_s}^2 + \sigma_{e_s}^2} = \mathrm{h}_s^2. \quad [22]$$

A simulation study was conducted to quantify the effects of phenotypic distribution shifts on genetic evaluations. Simulated stature records were generated for 5,141 animals, drawn from a subset of BS cattle based on a real pedigree containing 12,909 unique individuals. The pedigree was truncated to 3 generations. Two alternative population means were used: 147.32 cm (58 in) and 142.24 cm (56 in). An SD of 4.75 cm (1.87 in) was applied, ensuring a binning width comparable to that used by the BSA. The mean of 147.32 cm represents the newly updated BSA stature scoring system (denoted as EI_1), whereas 142.24 cm represents the mean of the previous scoring system (denoted as EI_0). For simplicity, the overall mean was treated as the sole fixed effect, whereas the random effects included additive genetic values and residuals. Heritability was set to 0.43, as reported by Wiggans et al. (2004).

The 1–9 scores were subsequently rescaled to a 1–50 range by multiplying by c = 5, consistent with the evaluation scale used by the Council on Dairy Cattle Breeding (CDCB, USA) for BS cattle. Genetic evaluations were performed using an animal model under various scenarios: M0 (the baseline model): the continuous stature measurements as phenotypes; M1A (Δ = 0), group means on the continuous scale as phenotypes; M1B (Δ = 0), 5× 1–9 scores derived without distribution as phenotypes; M2 (Δ = 2), 5× 1–9 scores with thresholds downscaled by Δ = 2 in as phenotypes; M3 (Δ = 4), 5× 1–9 scores with thresholds downscaled by Δ = 4 in as phenotypes. Accordingly, models M2 and M3 represent the old ordinal scoring systems, which are ineffectively updated (Δ = 2) or not updated at all (Δ = 4).

Model parameters were estimated using Bayesian inference via Markov chain Monte Carlo, assuming a flat prior for the overall mean and scaled inverse chi-squared priors for genetic and residual variances (Sorensen and Gianola, 2002). To minimize Monte Carlo error, each model was replicated 10 times, and the posterior means were averaged across replicates.

The baseline model M0 yielded a heritability estimate of h^2 = 0.433 (Table 1), closely matching the simulated value (0.43). When the ordinal scores were generated using thresholds consistent with the current phenotypic distribution, the estimated heritability (h^2 = 0.420–0.421) remained roughly consistent with the simulated heritability, regardless of whether the population means or ordinal scores were used as the phenotypes. However, when the scoring system was incorrectly or not updated, as in the cases of M2B (Δ = 2) and M3B (Δ = 4), the analyses produced systematically underestimated variance components compared with those from M1B (Table 1). In relative terms, the estimated genetic variance contracted more sharply than the estimated residual variance, reducing heritability to h^2 = 0.409 (M2B) and h^2 = 0.339 (M3B), respectively.

In what follows, we illustrate how to predict the impacts of phenotypic shifts using BLP. For M0: σ_y^2 = 22.7 cm^2, σ_u^2 = 10.0 cm^2, σ_e^2 = 13.1 cm^2, and $\mathrm{h}^2 = \frac{10.0}{10.0 + 13.1} \approx 0.433$ (Table 1). Using the exact category probabilities, the variance of the 1–9 scores is

$$\mathrm{Var}\left(s\right) = \sum_{j=1}^{9} j^2 p_j - \left(\sum_{j=1}^{9} j p_j\right)^2 = 28.39 - 5^2 \approx 3.39.$$

The BLP slope of 1–9 scores on genetic values (per cm) is

$$\beta = \frac{1}{\sigma_y} \sum_{j=1}^{k-1} \phi\left(\frac{\tau_j - \mu}{\sigma_y}\right) = \frac{1.82}{4.75} \approx 0.383.$$

The projected variances on the 1–9 scale are then given using Equations 15 and 16 as $\sigma_{u_s}^2 = \beta^2 \sigma_u^2 = 0.383^2 \times 10.0 \approx 1.47$; $\sigma_{e_s}^2 = \mathrm{Var}\left(s\right) - \beta^2 \sigma_u^2 = 3.39 - 1.47 \approx 1.92$ (Table 2).

According to Equations 19 to 21, rescaling the 1–9 scores to 5× (1–9) scores yields results comparable to those for M1B_Δ = 0 (Table 1): $\mathrm{Var}\left(5s\right) = 5^2 \times 3.39 \approx 84.7$; $\sigma_{u_{5s}}^2 = 5^2 \times 1.47 = 36.8$; $\sigma_{e_{5s}}^2 = 5^2 \times 1.927 = 48.2$. The projected heritability is

$$\mathrm{h}_{5s}^2 = \frac{36.8}{36.8 + 48.0} \approx 0.434.$$

Table 1. Variance components, heritability estimates, and rank correlations obtained under various scenarios[1,2,3]

Model	σ_y^2	σ_u^2	σ_e^2	S_u^2	S_e^2	h^2	ρ_y	ρ_u
M0	22.7 (0.250)	10.0 (0.846)	13.1 (0.708)	4.91 (0.620)	8.42 (0.826)	0.433 (0.033)	1 (0)	0.717 (0.008)
M1A: Δ = 0	22.0 (0.254)	9.44 (0.951)	13.0 (0.798)	4.42 (0.728)	11.6 (0.788)	0.421 (0.039)	0.985 (0.001)	0.710 (0.007)
M1B: Δ = 0	84.4 (0.489)	36.2 (3.05)	50.0 (2.94)	13.7 (1.97)	5.53 (0.764)	0.420 (0.035)	0.985 (0.001)	0.710 (0.007)
M2: Δ = 2	70.0 (0.622)	29.2 (2.78)	42.2 (2.49)	17.3 (2.19)	4.58 (0.611)	0.409 (0.037)	0.963 (0.001)	0.698 (0.007)
M3: Δ = 4	31.5 (0.509)	10.9 (1.23)	21.3 (1.04)	4.53 (0.648)	9.10 (0.937)	0.339 (0.036)	0.841 (0.003)	0.638 (0.009)

[1] σ_y^2 = phenotypic variance component on the measurement or ordinal scale; σ_u^2 = additive genetic variance; σ_e^2 = residual variance; S_u^2 = sample variance of estimated additive genetic values; S_e^2 = sample variance of residual effects; $\mathrm{h}^2 = \dfrac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$ = heritability.

[2] ρ_y = rank correlation between simulated phenotypes and the continuous or ordinal phenotypes used by each model; ρ_u = rank correlation between simulated additive genetic values and estimated additive genetic values under each model.

[3] M0 = model fitted on the original measurements as the phenotypes; M1A = model fitted on group means on the continuous scale as the phenotypes; M1B = model fitted on 5× 1–9 scores as the phenotypes without threshold shifts; M2 = model fitted on 5× 1–9 scores as the phenotypes after downscaling the thresholds by Δ = 2 in; M3 = model fitted on 5× 1–9 scores as the phenotypes after downscaling the thresholds by Δ = 4 in.

Interestingly, with a correctly updated scoring system, the linear projection yielded comparable results to the linear model fitted on the original scale.

Now, downscale the thresholds by Δ = 2 in (5.08 cm; Table 2). The variance on the 1–9 scale is $\mathrm{Var}\left(s^*\right) = 49.98 - 6.869^2 \approx 2.80$. The BLP slope becomes

$$\beta^* = \frac{1}{\sigma_y}\sum\nolimits_{j=1}^{k-1}\phi\left(\frac{\tau_j - \mu - \delta}{\sigma_y}\right) = \frac{1.608}{4.75} \approx 0.339.$$

Projected genetic and residual variances are $\sigma^2_{u_{s^*}} = \beta^{*2}\sigma_u^2 = 0.339^2 \times 10.0 \approx 1.15$ and $\sigma^2_{e_{s^*}} = \mathrm{Var}\left(s^*\right) - \beta^{*2}\sigma_u^2 = 2.798 - 1.154 \approx 1.64$.

On the rescaled 1–50 scale: $\mathrm{Var}\left(5s^*\right) = 5^2 \times \mathrm{Var}\left(s^*\right) = 67.0$, $\sigma^2_{u_{5s}} = 5^2 \times 1.154 = 28.9$, $\sigma^2_{e_{5s^*}} = 5^2 \times 1.64 = 41.0$, and $\mathrm{h}^2_{5s^*} = \dfrac{28.9}{28.9 + 41.0} \approx 0.412$.

These values closely matched those for model 2 (Δ = 2; Table 1). Essentially, the reduced heritability estimate reflected the impact of a shift in ordinal phenotype distribution.The results for downscaling thresholds further by Δ = 4 in (10.16 cm) were obtained in a similar way, where $\sum_{j=1}^{k-1}\phi\left(\dfrac{\tau_j - \mu - \delta}{\sigma_y}\right) = 0.934$ and $\beta^* = \dfrac{0.934}{4.75} \approx 0.197$. Shifting the thresholds downward by 4 in further increases the mass in the upper bins, thereby reducing the BLP slope, genetic variance, and heritability (Table 2).

Finally, a reduction in heritability inevitably entails a loss of selection accuracy. Linearly back-transforming EBVs from the ordinal scale to the latent continuous scale, the expected accuracy of selection under a shifted system (Δ > 0) can be expressed as

$$r\left(\Delta\right) = \sqrt{\mathrm{h}_s^2\left(\Delta\right)}. \qquad [23]$$

In our example, selection accuracy decreased from 0.659 (M0) to 0.650 (M1B: Δ = 0), 0.640 (M2: Δ = 2), and 0.582 (M3: Δ = 4). The rank correlations between simulated and estimated additive genetic values were slightly higher than $r(\Delta)$ but followed a similar decreasing trend: 0.717 (M0), 0.710 (Δ = 0), 0.698 (Δ = 2), and 0.638 (Δ = 4; Table 2).

For selecting the top-α animals, the expected response is

$$R = i_\alpha r\left(\Delta\right)\sigma_u, \qquad [24]$$

where i_α is the selection intensity and σ_u is the additive genetic standard deviation. For example, when selecting the top 5% (α = 0.05) of animals, $i_\alpha = \phi\left(z_{1-\alpha}\right)\big/\alpha$ with $z_{0.95} = 1.645$, which gives $i_{0.05} = 2.063$. The expected selection responses (in the same unit as the latent scale) are 1.35 (SD = 4.29) under Δ = 0, 1.32 (SD = 4.19) under Δ = 2, and 1.15 (SD = 3.65) under Δ = 4.

Table 2. Projected variance components and heritability estimates using a best linear projection approach[1]

		1–9 scores			5× 1–9 scores			
Δ	$\hat{\beta}$	σ_s^2	σ_u^2	σ_e^2	σ_s^2	σ_u^2	σ_e^2	h^2
0 in (0 cm)	0.382	3.39	1.47	1.92	84.8	36.8	48.0	0.434
2 in (5.08 cm)	0.339	2.80	1.15	1.64	67.0	28.9	41.0	0.412
4 in (10.16 cm)	0.197	1.25	0.39	0.857	31.2	9.75	21.4	0.313

[1] Δ = downscale offset of binning thresholds; $\hat{\beta}$ = regression coefficient; σ_s^2 = variance of 1–9 (or 5× 1–9) ordinal scores; σ_u^2 = projected additive genetic variance on the ordinal 1–9 (or 5× 1–9) scale; σ_e^2 = projected residual variance on the ordinal 1–9 (or 5× 1–9) scale; $\hat{\mathrm{h}}^2$ = heritability estimate.

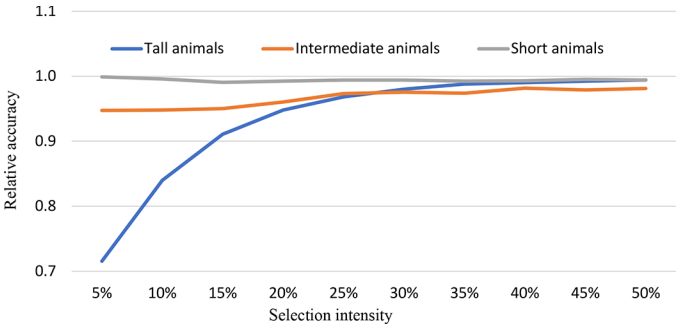

Figure 1. Selection accuracy using 5× 1–9 ordinal scores as phenotypes under 3 selection scenarios, relative to the selection accuracy using original stature measurements as phenotypes.

In addition to overall accuracy, we also assessed accuracy for specific selection objectives—selecting the tallest, intermediate, and smallest animals—under selection intensities ranging from 5% to 50%. The relative selection accuracy (**RSA**) was defined as the ratio of rank correlation between the estimated additive genetic values obtained from the ordinal-scale evaluation and those on the continuous-scale baseline model (M0). This metric reflects a realistic breeding scenario in which actual breeding values are unknown, and evaluations based on observed phenotypes serve as the benchmark, against which alternative scoring systems are compared.

Across 10 replicates, selecting the top 5% of animals based on outdated scores ($\Delta = 2$) resulted in an average RSA reduction of 28.5% (Figure 1). This substantial loss in selection accuracy stemmed from misclassifying nonelite animals as "very tall" due to score compression and misaligned bin boundaries, which concentrated scores in the highest categories (see also the bottom-left panel of the Graphical Abstract). Nevertheless, the adverse effect diminished as selection intensity broadened: the RSA loss declined from 28.5% to 16.0% and then to only 0.6% as selection proportion increased from 5% to 10% and then to 50%. This dilution effect arises because a broader selection included animals less affected by threshold misalignment. Hence, outdated scoring thresholds can systematically erode both selection accuracy and expected genetic gain, with the greatest impact observed under intense selection for extreme phenotypes.

In conclusion, phenotypic shifts have tangible and significant consequences for genetic evaluations and selection outcomes. Outdated scoring systems distort the mapping between biological measurements and categorical scores, leading to biased variance components, underestimated heritability, and reduced selection accuracy. Therefore, periodic recalibration of threshold boundaries is essential to keep the ordinal scoring systems aligned with current population distributions and to safeguard the validity of genetic improvement programs.

While stature provides a clear example of scale drift in an ordinal scoring system, other traits recorded on discrete scales—such as BCS—may also be susceptible to similar distributional shifts over time. A systematic investigation of these traits could help determine whether recalibration is needed more broadly across fitness and welfare-related phenotypes.

References

Casella, G., and R. L. Berger. 2002. Statistical Inference. 2nd ed. Duxbury Press, Pacific Grove, CA.

Manafiazar, G., L. Goonewardene, F. Miglior, D. H. Crews Jr., J. A. Basarab, E. Okine, and Z. Wang. 2016. Genetic and phenotypic correlations among feed efficiency, production and selected conformation traits in dairy cows. Animal 10:381–389. https://doi.org/10.1017/S1751731115002281.

Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Springer-Verlag, New York, NY.

Stein, C. 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Pages 197–206 in Proc. 3rd Berkeley Symp. Math. Stat. Probab. University of California Press, Berkeley, CA.

VanRaden, P. M., and G. R. Wiggans. 1995. Productive life evaluations: Calculation, accuracy, and economic value. J. Dairy Sci. 78:631–638. https://doi.org/10.3168/jds.S0022-0302(95)76674-7.

Veerkamp, R. F., and S. Brotherstone. 1997. Genetic correlations between linear type traits, food intake, live weight and condition score in Holstein Friesian dairy cattle. Anim. Sci. 64:385–392. https://doi.org/10.1017/S1357729800015976.

Visscher, P. M., and M. E. Goddard. 2011. Cattle gain stature. Nat. Genet. 43:397–398. https://doi.org/10.1038/ng.819.

Wiggans, G. R., N. Gengler, and J. R. Wright. 2004. Type trait (co)variance components for five dairy breeds. J. Dairy Sci. 87:2324–2330. https://doi.org/10.3168/jds.S0022-0302(04)70054-5.

Notes

Xiao-Lin Wu, https://orcid.org/0000-0002-5604-9220
Allicia Horn, https://orcid.org/0009-0001-2193-6225
Ezequiel Nicolazzi, https://orcid.org/0000-0001-9680-5054
John B. Cole https://orcid.org/0000-0003-1242-4401

This study received no external funding.

No human or animal subjects were used, so this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board.

The authors have not stated any conflicts of interest.

Nonstandard abbreviations used: BLP = best linear projection; BS = Brown Swiss; BSA = Brown Swiss Association; EI = equal interval; PDF = probability density function; RSA = relative selection accuracy.