



## Correlation and path analysis of morphological traits in a bread wheat collection using Structural Equation Modelling

**Nargiz Eyyubova\***

Postgraduate Student

Genetic Resources Institute of the Azerbaijan National Academy of Sciences

AZ1106, 155 Azadlıg Ave., Baku, Azerbaijan

<https://orcid.org/0009-0008-6615-7234>

**Hamlet Sadiqov**

Doctor of Biological Sciences, Associate Professor

Genetic Resources Institute of the Azerbaijan National Academy of Sciences

AZ1106, 155 Azadlıg Ave., Baku, Azerbaijan

<https://orcid.org/0009-0006-9302-6817>

**Sevil Sadigova**

PhD in Biological Sciences, Leading Researcher

Genetic Resources Institute of the Azerbaijan National Academy of Sciences

AZ1106, 155 Azadlıg Ave., Baku, Azerbaijan

<https://orcid.org/0009-0009-4718-9561>

**Nicat Gasim**

Assistant Professor

Karabakh University

AZ2600, Khankendi, Azerbaijan

<https://orcid.org/0000-0002-3976-1283>

**Abstract.** The aim of this study was to investigate the structural determinants of grain yield per plant in bread wheat (*Triticum aestivum* L.) and to assess the contributions of morphological and yield-related traits. Descriptive statistics, Pearson's correlation analysis, path coefficient analysis, and Structural Equation Modelling (SEM) were employed. A total of 53 genotypes from the bread wheat collection of the Genetic Resources Institute (Baku, Azerbaijan) were evaluated for grain weight per spike, thousand-kernel weight (TKW), grains per spike (GPS), plant height (PH), and main spike length (SL). Analysis revealed that grain yield per plant (GWPP) was primarily influenced by numerical (GPS,  $r = 0.47$ ) and mass-related (TKW,  $r = 0.29$ ) traits, whereas architectural traits (PH, SL) contributed mainly indirectly. SEM showed that grain weight per spike (GWPS) acted as a central mediating variable, channelling the effects of GPS and TKW to overall yield. The mass/sink axis (TKW  $\rightarrow$  GWPS) exhibited a stronger direct effect (standardised coefficient = 2.18) than the architectural axis (0.41), while architectural traits primarily supported sink capacity indirectly. The final model demonstrated excellent fit ( $\chi^2(3) = 1.26$ ,  $p = 0.74$ ; Comparative Fit Index = 1.00; Root Mean Square Error of Approximation = 0.00; Standardised Root Mean Square Residual = 0.036), supporting the biological interpretability of the observed trait relationships. These results provide practical insights for plant breeders and geneticists seeking to select high-yielding wheat genotypes under

### Suggested Citation:

Eyyubova, N., Sadiqov, H., Sadigova, S., & Gasim, N. (2026). Correlation and path analysis of morphological traits in a bread wheat collection using Structural Equation Modelling. *Biological Systems: Theory and Innovation*, 17(1), 78-90. doi: 10.31548/biologiya/1.2026.78.

\*Corresponding author





controlled field conditions and can be utilised in wheat breeding programmes to optimise strategies for improving bread wheat production

**Keywords:** *Triticum aestivum* L.; grain yield; classical framework; chemical treatments; source-sink dynamics

## INTRODUCTION

Investigating the correlation and path relationships of morphological traits in bread wheat using Structural Equation Modelling (SEM) is essential for understanding the complex interactions among yield components, including grain number, grain weight, and plant architecture. Such analysis enables the identification of direct and indirect effects of mass-related, numerical, and architectural traits on overall productivity, providing insights into which characteristics most strongly influence yield formation. By quantifying these relationships within diverse wheat genotypes, breeders can prioritise traits that contribute most effectively to grain yield, optimise selection strategies, and improve the efficiency of breeding programmes. Additionally, SEM allows for the integration of multiple traits into a coherent model, revealing hidden patterns and mediation effects that traditional correlation analyses may overlook, thereby supporting data-driven decision-making in wheat improvement under specific environmental and agronomic conditions. The theoretical foundation of this study rests on two complementary axes.

The first is the numerical axis, represented mainly by the number of grains per spike (GPS). L. Vicentin *et al.* (2024) investigated the contribution of the numerical component to yield variability in bread wheat and established that a large proportion of yield differences originated from this axis. T. Fischer *et al.* (2024) examined the impact of numerical traits on overall grain production and highlighted their critical role in genotype selection. K. Wang *et al.* (2023) analysed the effects of limited photosynthetic supply on mass per sink capacity and demonstrated its importance for maintaining yield stability under stress conditions. In research C. Zhang *et al.* (2022) explored the physiological processes underlying average grain mass (TWK), including carbon allocation, assimilate transport, and endosperm development during grain filling. R.B. Flavell (2023) studied the architectural axis and its regulation of source-sink balance, showing how synergistic operation

of numerical and architectural traits enhanced grain weight per spike (GWPS) and grain yield per plant (GWPP). X.-L. Wu *et al.* (2022) interpreted yield formation within the classical source-sink framework, emphasising the equilibrium between assimilate supply and sink capacity during pre- and post-flowering phases. In particular, T.O. Milkessa (2022) demonstrated that grain yield was frequently limited by sink capacity during grain filling, indicating the necessity to improve both number-related (GPS, NoSPS) and mass-related traits (TWK) simultaneously to enhance overall GWPP through the mediating effect of GWPS. In the study, T.H. Borkhatariya *et al.* (2025) investigated the relationships between yield and yield-related traits in bread wheat using correlation and path coefficient analysis. Furthermore, they identified key traits, such as biological yield and number of effective tillers per plant, which had significant direct effects on grain yield. In addition, they evaluated the influence of plant height and spike length on overall productivity, determining their indirect effects through key yield components. Moreover, the study emphasised the importance of selecting genotypes with optimised combinations of numerical and mass-related traits to enhance stable grain yield under diverse environmental conditions. The study identified gaps including limited genotype coverage, insufficient multifactor analysis of numerical and mass-related yield components, inadequate evaluation of architectural traits under diverse environments, lack of multi-season and multi-location trials, and insufficient integration of modern Structural Equation Modelling (SEM) approaches, restricting the predictive capacity for yield stability and the selection of high-performing bread wheat cultivars.

The objective was to examine the factors influencing GWPP in bread wheat (*Triticum aestivum* L.) and to assess the roles of various morphological and yield-associated traits. Accordingly, in this study, biologically justified alternative SEM pathway models were constructed, evaluated through



indices such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardised Root Mean Square Residual (SRMR), and interpreted based on standardised coefficients and bootstrap validation. Given the single-environment nature of the dataset, the objective was to explore the internal structural relationships among yield components rather than to make broad, environment-general breeding recommendations.

## MATERIALS AND METHODS

The experiment was conducted during the 2025 spring growing season. The experiment was carried out in the experimental field of the Azerbaijan State Genetic Resources Institute, which is officially designated for scientific research on cereal crops. A total of 53 bread wheat varieties 'Oyxan', 'Ekinchi-84', 'Tereqqi', 'Murov', 'Qaudio', 'Mirbeshir-128', 'Romanna', 'Ferehim', 'Midas', 'Nogal', 'Yasavul', 'Olimp', 'Nurlu-99', 'Khamsa', 'Qualite', 'Deabar', 'Chempion', 'Bayraqdar', 'Azeri', 'Qobustan', 'Onur', 'Yuka', 'Lider', 'Shefeq-2', 'Mirastar', 'Fatime', 'Shahbugda', 'Xlebodarka', 'Farandol', 'Khazri', 'Renan', 'Atilla', 'Batko', 'Ugur', 'Baqrat', 'Auradur', 'Leyaqetli-80', 'Guneshli', 'Qizilbughda', 'Vusal', 'Metin', 'Kroshka', 'Zirve-85', 'Esgeran', 'Stan-Krasnodar', 'Gallio', 'Murov-2', 'Balaton', 'Seba-2', 'Qirmizigul', 'Tunc', 'Maruzio', 'Gyrmyzy bughda' were used as experimental material. Each genotype was grown in a single-row plot with a single replication (one plot per genotype) under uniform management. The field was prepared using conventional tillage operations applied uniformly across the entire experimental area prior to sowing. All genotypes were sown on the same date, at the same seeding rate, row spacing, and planting depth to eliminate variation in plant density and early establishment. Basal mineral fertilisation was applied evenly across the field according to regional agronomic recommendations for bread wheat, ensuring equal nutrient availability for all plots. Irrigation was carried out uniformly across the experimental site based on crop water requirements and prevailing climatic conditions, preventing differential moisture stress among genotypes.

Weed control was maintained through identical mechanical and/or chemical treatments applied simultaneously to the whole field area. Preventive plant protection measures against common wheat pests and diseases were

implemented uniformly, and no genotype-specific chemical treatments were applied. No thinning or selective removal of weaker plants was performed within individual plots, in order to avoid artificial alteration of genotype performance. All plots were managed under the same environmental and agronomic conditions without differential interventions. These standardised practices ensured that observed differences among genotypes reflected genetic variability rather than variation in management conditions. This experiment represents a single-environment evaluation (one growing season and one location). The un-replicated (single-replication) structure and lack of multi-environment testing limit the generalisability of estimated associations; thus, the SEM was used primarily to explore internal trait relationships within the studied germplasm set.

The study was conducted in full compliance with the ethical principles of the Convention on Biological Diversity (1992), and all established ethical requirements were strictly observed. The plant materials used in this study were part of the official wheat germplasm collection of the Genetic Resources Institute, and no special ethical approval was required. Descriptive statistics (mean, standard deviation, and coefficient of variation) were computed for all morphological and yield-related traits. Pearson's correlation coefficients were calculated to quantify pairwise associations among traits. To determine the magnitude and direction of both direct and indirect effects of the studied traits on GWPP, path analysis was performed following the classical framework proposed by S. Wright (1921) and further developed by C.C. Li (1975). Given the moderate sample size ( $n = 53$ ), particular emphasis was placed on model parsimony and biological interpretability. The SEM was therefore specified with a limited number of latent constructs and theoretically justified pathways to reduce the risk of overparameterisation. Model fit indices were interpreted cautiously, recognising that very high values (e.g., CFI  $\approx 1.00$ , RMSEA  $\approx 0.00$ ) may partly reflect low degrees of freedom rather than perfect biological representation. All statistical computations, including descriptive statistics, correlation analysis, and path coefficient estimation, were carried out using the Jamovi software (version 2.4.14) and Python (version 3.12.4) environments.

## RESULTS

First, all candidate variables were subjected to a preliminary examination using descriptive statistics and a correlation matrix. Subsequently, a biologically justified multiple path model (MPM) was constructed, and comparisons were performed based on fit indices ( $\chi^2$ , CFI, TLI, RMSEA, SRMR), parsimony indicators (Parsimony Normed Fit Index (PNFI), Akaike Information Criterion/Bayesian Information Criterion (AIC/BIC)), and standardised coefficients/Bootstrap GA-based comparisons. During this screening process, the variables NoPT (number of productive tillers), NoSPS (number of spikelets per spike), and GPS (number of grains per spikelet) were evaluated and tested in structural blocks in different schemes. However, when making the final (best fit-simplest structure) selection, adding these variables to the model as free paths (i) did not increase global fit, (ii) did not independently and consistently increase the explained variance contribution compared to the basic components already present in the model (especially GWPS and TWK), and (iii) weakened parsimony in the context of the sample size ( $n = 53$ ). The selected final model provides excellent global fit and high interpretability by

focusing on the relationship between mass/sink capacity captured through the information content GWPS and TWK and GWPP. This choice is based on two principles: (i) The principle of mediation – the effect of numerical components (NoSPS, GPS) is primarily reflected in yield through the GWPS channel; therefore, repeatedly introducing the same information into the model both directly and indirectly creates unnecessary freedom and uncertainty. (ii) The principle of parsimony – removing paths that do not increase fit in a small/medium sample ( $n = 53$ ) increases the stability of the coefficients and the reliability of the inferences. However, the descriptive statistics and bivariate relationships of the NoPT-NoSPS-GPS variables are included in the report; their roles in the theoretical framework (representing numerical capacity and spike architecture) are discussed in the discussion section. Thus, while theoretical integrity is preserved, the empirically strongest and simplest structure is carried over to the results section. The descriptive statistics obtained at the beginning of the study in Table 1 reveal both the central tendencies and the distributional patterns of the variables, guiding how subsequent modelling steps should be designed.

**Table 1.** Descriptive statistics

	GWPP	TKW	GPS	PH	LoS
Mean	6.84	49.3	45.7	103.6	11.3
Median	6.58	49	45.4	102.8	10.3
Maximum	20.3	62	70.8	126	38.9
Minimum	2.60	40	32	83.4	6.8
Std. Dev.	2.79	5.38	7.50	9.50	5.03
Coef.Var	40.8	10.9	16.4	9.16	44.5
Skewness	2.19	0.28	0.72	0.07	4.14
Kurtosis	11.3	2.89	4.07	2.59	20.8
Jarque-Bera	192.6	0.724	7.074	0.413	855.1
Probability	0.000	0.696	0.029	0.813	0.000
Observations	53	53	53	53	53

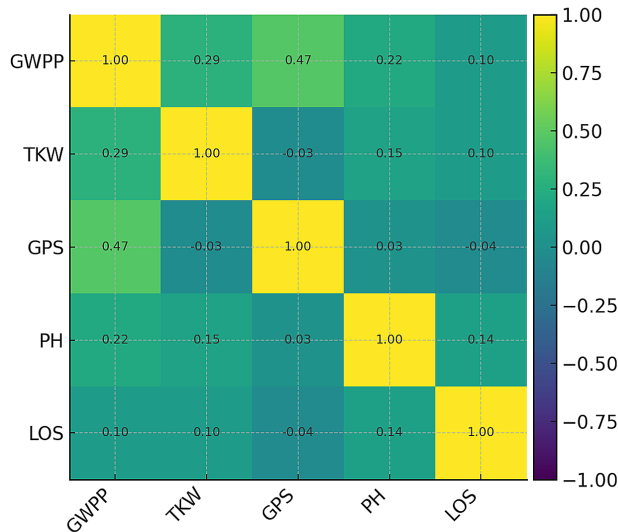
**Source:** compiled by the authors

GWPP: The mean is 6.84 g plant<sup>-1</sup>, the median is 6.58, and the range is 2.60-20.30. With a standard deviation of 2.79 and CV of 40.8%, a high dispersion is observed. The skewness = 2.19 and kurtosis = 11.3 measures indicate that the distribution is distinctly right-skewed and heavy-tailed. The Jarque-Bera = 192.6 ( $p < 0.001$ ) result confirms a significant deviation from normal distribution.

TKW: The mean is determined to be 49.3 g 1,000 grains<sup>-1</sup>, the median is 49.0, and the range is 40-62. Standard deviation = 5.38 and CV = 10.9% indicate low variability. Skewness = 0.28 and kurtosis = 2.89 values show that the distribution is close to normal; Jarque-Bera = 0.724 ( $p = 0.696$ ) confirms that this fit is statistically supported. GPS: The mean is 45.7, the median is 45.4, and

the range is 32-70.8. With a standard deviation of 7.50 and a CV of 16.4%, a moderate level of dispersion is observed. Skewness = 0.72 indicates a moderately positive tail, and kurtosis = 4.07 indicates slight peakedness. The Jarque-Bera = 7.074 ( $p=0.029$ ) value indicates a limited deviation from the normal form. PH: The mean is 103.6 cm, the median is 102.8 cm, and the range is 83.4-126 cm. With a standard deviation of 9.50 and CV = 9.16%, the lowest dispersion is observed in this variable. The measures skewness = 0.07 and kurtosis = 2.59 indicate that the distribution is symmetric and close to normal; Jarque-Bera = 0.413 ( $p = 0.813$ ) confirms that the fit to the normal distribution is

acceptable. LOS (ear length). The mean is recorded as 11.3 cm, the median as 10.3 cm, and the range as 6.8-38.9 cm. With a standard deviation of 5.03 and a CV of 44.5%, the highest heterogeneity is observed in this variable. Skewness = 4.14 and kurtosis = 20.8 indicate a pronounced right-tail and heavy-tail structure. The Jarque-Bera = 855.1 ( $p < 0.001$ ) result confirms a strong deviation from normal distribution. Following descriptive statistics, bivariate relationships between variables were examined using the Pearson correlation matrix in Figure 1, where coefficient magnitudes are categorised according to conventional thresholds ( $\approx 0.10$ - $0.29$  weak;  $\approx 0.30$ - $0.49$  moderate;  $\geq 0.50$  strong).



**Figure 1.** Correlation matrix heatmap

**Source:** compiled by the authors

GWPP was found to have a moderate and positive relationship with GPS ( $r=0.471$ ). This result suggests that yield is significantly influenced by the numerical component (number of grains per head). A weak-moderate positive relationship was found with TKW ( $r=0.286$ ); it is understood that the increase in average grain mass contributes to yield, but to a more limited extent compared to GPS. Relationships with architectural indicators are weak: PH-GWPP  $r=0.218$ , LoS-GWPP  $r=0.102$ ; this pattern implies that architecture reflects on yield more through indirect channels. It has been determined that the relationship between TKW and GPS is negligible ( $r = -0.028$ ). This finding

indicates that no significant “compensation/trade-off” signal has emerged between the numerical and mass components under these materials and conditions, suggesting that the two components can move independently. The weak positive relationships between TKW and PH ( $r=0.153$ ) and LoS ( $r=0.103$ ) suggest that mass capacity is only moderately correlated with architecture. GPS has a very low correlation with architectural indicators (PH-GPS  $r = 0.025$ , LoS-GPS  $r = -0.037$ ). This indicates that the processes determining spikelet/floret dynamics and fertility can operate independently of the measured architectural indicators. On the other hand, GPS’s moderate positive correlation

with GWPP ( $r=0.471$ ) confirms its effect on yield. A weak positive relationship was found between the two architectural traits ( $r=0.139$ ). This magnitude suggests that architectural components, while changing together, are not powerful enough to explain yield variation on their own and should be considered as control or indirect effects in structural models. No high correlation ( $|r| \geq 0.85$ ) was found between predictors. In particular, the fact that the TWK-GPS relationship is close to 0 indicates that the risk of multicollinearity is low when both numerical and mass components are included in the model simultaneously. This profile statistically supports the separation of the effects of the mass axis (TWK/GWPS) and the numerical axis (GPS) in road analysis. The pattern in the

matrix shows that productivity (GWPP) is primarily driven by numerical (GPS) and mass (TWK) components, while architectural indicators (PH, LoS) exhibit weak singular relationships. Therefore, it is appropriate to model the yield flow in the final structural design along the GWPS/TKW axis and based on the assumption of the indirect effect of GPS; architectural variables should be considered in an indirect/conditional role when there is a theoretical justification. Following descriptive statistics and correlation patterns, a Structural Equation Model was estimated to evaluate the extent to which the theoretical flow (architecture  $\rightarrow$  numerical/mass components  $\rightarrow$  ear mass  $\rightarrow$  yield) was supported by the data. The goodness-of-fit results for this model are presented in Table 2.

**Table 2.** Model fit indices for the estimated SEM structure

Model tests			
Label	$\chi^2$	df	p
User model	1.26	3	0.740
Baseline model	25.08	10	0.005

**Source:** compiled by the authors

The chi-square test for the user model,  $\chi^2(3) = 1.26$ ,  $p = 0.740$ , indicates that there is no significant difference between the observed covariance matrix and the covariance predicted by the model. This result indicates that the absolute fit is very good and that the model contains at least 3 over-specified constraints (parsimony). The baseline model provided for comparison was significantly rejected with  $\chi^2(10) = 25.08$ ,  $p = 0.005$ ; the data does not support the assumption that the variables are independent of each other. This dual table also implies that the relative fit indices (CFI, TLI/Non-Normed Fit Index (NNFI), Bollen's Incremental Fit Index (IFI),

Normed Fit Index (NFI), Relative Fit Index (RFI)) will be high; indeed, CFI reaches 1.00 as required by the formula (when  $\chi^2_{\text{user}} - df_{\text{user}} \leq 0$ , the numerator becomes 0). Furthermore, because  $\chi^2$  is smaller than df, Root Mean Square Error of Approximation (RMSEA) is expected to be very close to 0. In summary, the model tests results show that the established structure explains the data successfully and parsimoniously; therefore, in the next step, the standardised path coefficients and indirect/total effects can be interpreted with confidence. The detailed fit indices, including SRMR and RMSEA values, are summarised in Table 3.

**Table 3.** Fit indices

95% confidence intervals				
SRMR	RMSEA	Lower	Upper	RMSEA p
0.036	0.000	0.000	0.163	0.000

**Source:** compiled by the authors

The Standardised Root Mean Square Residual (SRMR) = 0.036 indicates that the average size of standardised residuals is low and that the  $\leq 0.08$  (preferably  $\leq 0.05$ ) threshold is clearly met; this implies that the model captures the covariance

structure at a detailed level. The RMSEA = 0.000 shows that the model error is well below the "close fit" threshold ( $\approx 0.05-0.06$ ), and the RMSEA  $p(\text{close}) = 0.777$  indicates that the "close fit exists" hypothesis cannot be rejected – strong evidence in

favour of a close fit. The lower bound of the RMSEA 95% CI [0.000, 0.163] touching 0 indicates that the point estimate is very low, while the width of the upper bound may reflect uncertainty due to the small df. When considered together, the SRMR and RMSEA results in Table 3 confirm that the model explains the data at a very good/close fit level, and that the parameter (path coefficient) interpretations can be made with confidence. It should be

noted that extremely high incremental fit indices (e.g., CFI = 1.00, RMSEA = 0.00) are not interpreted here as evidence of biological perfection. In small or moderately sized samples with low degrees of freedom, such values may arise from model parsimony and constrained structure. Accordingly, model evaluation focused on consistency with biological theory rather than on numerical maximisation of fit indices (Table 4).

**Table 4.** User model versus baseline model

Fit measure	Model	Reference/Cutoff	Assessment
CFI	1.000	≥0.95 good (≥0.90 acceptable)	Excellent
TLI	1.386	≥0.95 good	Excellent (interpreted as ≈1.00)
NNFI	1.386	≥0.95 good	Excellent (≈1.00)
RNI	1.116	≥0.95 good	Excellent (≈1.00)
NFI	0.950	≥0.95 good (≥0.90 acceptable)	Good (at threshold)
RFI	0.833	≥0.90 preferred	Moderate/Below preferred
IFI	1.079	≥0.95 good	Excellent (≈1.00)
PNFI	0.285	No hard cutoff; higher = more parsimonious (≈ >0.50 moderate)	Low parsimony (few df)

**Note:** RNI – Relative noncentrality index

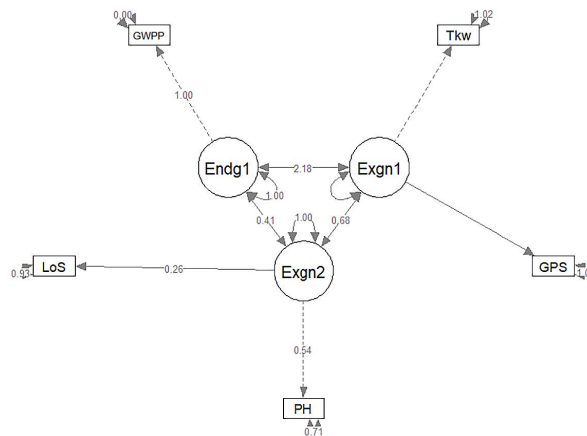
**Source:** compiled by the authors

After confirming absolute and local fit using chi-square, SRMR, and RMSEA, additional incremental (relative) fit indices were examined to evaluate how much the user model improved compared to the baseline model, as summarised in Table 4. The incremental/relative indices indicate that the model fits very well compared to the independence model: CFI = 1.000 is excellent; TLI/NNFI and IFI exceeding 1 are due to software rounding and are interpreted in practice as ≈1.00 (perfect fit). NFI = 0.950 fully meets the good threshold value. Although RFI = 0.833 is relatively low, this index is sensitive to df and can be explained by the small number of constraints. PNFI = 0.285 indicates low parsimony (few constraints/few df), which is not a fit issue but rather a measure of the model's simplicity. The coefficients in the path diagram (factor loadings, error terms, and direct/indirect effects) were interpreted on a numerical basis (Grace, 2009). It was determined that the latent structure of Endg1 was scaled with a loading of 1.00 using GWPP and therefore the error term for GWPP was 0.00. The connection of Exgn1, representing the mass/sink axis, with TWK was ≈ 1.00 and the indicator error term was at the level of  $\epsilon_{TKW} \approx 1.02$ . The

architectural/vigor axis Exgn2 was carried by PH with a loading of 0.54 ( $\epsilon_{PH} \approx 0.71$ ) and LoS with a loading of 0.26 ( $\epsilon_{LoS} \approx 0.93$ ). This pattern indicates that PH is more strongly represented on the architectural axis than LoS (0.54 > 0.26) and that the relative error share in the LoS measurement is higher (0.93 > 0.71). For the numerical efficiency indicator at the head level, an error term  $\epsilon_{GPS} \approx 1.05$  was recorded; it was also observed that GPS has out-of-model determinants (Ochagavía *et al.*, 2018; Slafer *et al.*, 2023; Wang *et al.*, 2023). Overall, these results confirm that the model exhibits strong relative fit, justifying the interpretation of the standardised path coefficients with confidence. The final structural model, depicting the direct and indirect relationships among yield-related traits, is presented in Figure 2. The strong association between the mass/sink axis and yield (Exgn1 → Endg1 = 2.18) indicates that, within this dataset, maintaining kernel weight during the grain-filling period represents a dominant structural component of yield formation. In this single-environment evaluation, the observed dominance of the mass/sink-related axis is consistent with the recognised importance of kernel weight maintenance. However,

specific agronomic recommendations (e.g., irrigation management or late-season nitrogen strategies) cannot be inferred from the present design and should be explicitly tested in multi-environment and/or managed-stress experiments. The positivity of  $\text{Exgn1} \rightarrow \text{GPS}$  conveys the message that suppressing TKW is not mandatory while increasing grain number; it is understood that the target of “ $\text{GPS}\uparrow + \text{TKW} \approx \text{stable}$ ” is achievable with appropriate genotype  $\times$  environment arrangements.  $\text{Exgn2} \rightarrow$  Based on the  $\text{Endg1} = 0.41$  result, the architecture should be maintained within a yield-supporting range: Considering the strong load of 0.54 for PH, sufficient vigour should be ensured without increasing the risk of lodging. Due to the weak load of 0.26 for LoS and the error  $\varepsilon_{\text{LoS}} \approx 0.93$ , monitoring spike length as a secondary/auxiliary indicator should be sufficient. Ultimately, the strategy that most rapidly reflects an increase in GWPP is “supporting GPS while preserving (and if possible increasing) TWK” and

managing architecture at a level that facilitates this process (Zhang *et al.*, 2020; Wu *et al.*, 2022; Fang *et al.*, 2022). From a methodological perspective, the moderate sample size necessitates cautious interpretation of the SEM results. Although the model was intentionally specified in a parsimonious manner, the possibility of capitalising on sample-specific covariance patterns cannot be entirely excluded. Accordingly, the SEM presented here should be regarded as an exploratory representation of yield-component structure and requires confirmation using larger and replicated datasets. The study is based on a single location, a single growing season, and one replication per genotype. Consequently, the magnitude of correlations and SEM pathways may differ under alternative environmental conditions, management regimes, or stress scenarios. The present findings should therefore be regarded as exploratory and hypothesis-generating, providing a basis for future validation in multi-environment studies.



**Figure 2.** Path diagram of the SEM depicting direct and indirect relationships among variables among key morphological and yield-related traits in bread wheat

Source: compiled by the authors

## DISCUSSION

In the results of the present study, it was determined that the main components influencing yield formation in the evaluated bread wheat genotypes were grain weight per spike and thousand-kernel weight, which had a strong direct positive effect on grain yield per plant, whereas architectural traits (plant height and spike length) acted as secondary factors influencing yield mainly

indirectly through numerical and mass components. These findings are consistent with M. Barman *et al.* (2020), who reported a significant contribution of morpho-physiological traits and grain Fe and Zn content to yield and its components, while A. Baye *et al.* (2020) highlighted the direct and indirect effects of genotypic and phenotypic traits on grain yield in advanced wheat lines, confirming the role of both mass and numerical

components in productivity. Furthermore, A.A. Ezici *et al.* (2022) demonstrated that morphological and quality characteristics determine yield variability, which aligns with the relationships observed in the current study between mass and grain number indicators. Overall, all three studies support that the direct effect of mass-related components on yield is dominant, whereas numerical and architectural traits play auxiliary or indirect roles. In the study, G. Kour *et al.* (2025) assessed genetic variability and applied correlation and path analysis to determine the direct and indirect influences of morphological traits on grain yield. Furthermore, they demonstrated that certain architectural and numerical traits contributed differently to overall productivity, providing insights for targeted selection. S. Mahpara *et al.* (2024) analysed diverse wheat genotypes to quantify the effects of spike length, plant height, and grain weight on grain yield. In addition, they established which traits could be most efficiently manipulated to enhance yield performance in breeding programmes. In the article, A.G. Rajput *et al.* (2025) conducted correlation and path coefficient analysis in bread wheat to identify the relative contributions of morphological traits to productivity. Furthermore, they reported that both direct and indirect pathways of certain traits significantly affected grain yield, highlighting the complexity of yield formation. M. Yadav *et al.* (2025) examined genetic parameters and performed correlation and path analysis for yield and yield-contributing traits. Moreover, they provided recommendations for breeders on prioritising specific traits to achieve higher and more stable wheat yields.

In the present study, the dominant contribution of mass-related components, specifically grain weight per spike and thousand-kernel weight, to overall yield was confirmed, with architectural traits such as plant height and spike length exerting mainly indirect effects through numerical and mass traits. These findings are consistent with M.I. Ullah *et al.* (2021), who reported that grain yield in bread wheat lines is primarily influenced by mass and numerical traits, and that direct and indirect correlations among yield components are critical for yield improvement. Similarly, I. Havryliuk & H. Kovalyshyna (2025) and S. Arshad *et al.* (2025) showed that in soft winter wheat hybrids, morphological traits of the spike and

productive components of the ear are strongly associated with grain yield, supporting the conclusion that mass and spike-related traits play a dominant role, while architectural features act indirectly. Collectively, these studies reinforce that maintaining mass-related traits is essential for yield optimisation, whereas numerical and architectural traits contribute in secondary or mediating roles.

The study demonstrated that grain yield formation in bread wheat is primarily governed by mass-related components, particularly GWPS and TKW, which exerted the strongest direct effects on GWPP. Numerical components, represented by GPS, and architectural traits, including PH and LoS, contributed predominantly through indirect pathways mediated via GWPS and TKW. Correlation and path analyses confirmed that mass and numerical traits operate largely independently, with minimal trade-off, enabling simultaneous improvement of both components. Specifically, V. Chauhan *et al.* (2025) examined seventy-five wheat genotypes and evaluated their phenotypic and genotypic performance for yield-related traits. In addition, they highlighted that seed yield per plant was strongly associated with number of grains per spike and thousand-kernel weight, providing guidance for superior cultivar selection. In the article, M. Javed *et al.* (2024) and G.P. Sharda *et al.* (2023) estimated heritability and genotypic variability among bread wheat genotypes and analysed correlations for yield-attributing traits. Moreover, they emphasised the traits that could be prioritised in breeding programmes to improve grain yield stability under varying conditions.

The results indicated that mass-related traits are the most influential factors in yield determination, while numerical and architectural characteristics play supporting roles. D. Zewdu *et al.* (2024) investigated the relationships among yield and yield component traits in bread wheat (*Triticum aestivum* L.) genotypes. They applied correlation and path analysis to identify key traits influencing grain yield and to quantify their direct and indirect effects. These findings highlighted the importance of prioritising breeding and agronomic strategies that enhance grain mass while maintaining spike fertility and optimal plant architecture to maximise productivity. The study provided a foundation for further research under multi-environment trials to validate the observed relationships and

explore genotype  $\times$  environment interactions affecting yield components.

## CONCLUSIONS

The study demonstrated that wheat yield formation is governed predominantly by the interaction of numerical and mass-related traits rather than by architectural morphology. Among the examined traits, the number of GPS and the TWK jointly determined the majority of yield variation, whereas plant height (PH) and ear length (LOS) exhibited only indirect contributions through their influence on spike vigour and assimilate allocation. The SEM confirmed that GWPS functions as the central mediating variable, transmitting the effects of GPS and TWK onto overall GWPP. The quantitative results clearly emphasised that the mass/sink axis (TWK + GWPS) exerts a substantially stronger direct effect on yield compared with the architectural axis, confirming that the maintenance of sink capacity during the grain-filling phase is critical for achieving high productivity. The positive association between the mass and numerical axes suggests that simultaneous improvement in both grain number and grain mass is attainable without a strong trade-off, provided that environmental and genetic conditions are properly balanced. From a breeding perspective, the observed structure suggests that, within this germplasm and under the studied conditions, GPS and TWK-related components merit attention as primary yield-associated traits. The weak direct associations of architectural traits imply that their role is largely indirect, acting through support of sink-related processes rather than as primary drivers of yield variation. Nevertheless, the extent to which these relationships hold across years, environments, and management regimes should be confirmed using replicated multi-location trials before agronomic or breeding recommendations are generalised. The structural relationships identified here should be interpreted as hypothesis-generating and provide a basis for future SEM analyses using larger, multi-environment datasets. Overall,

the study provided an integrative framework for understanding yield architecture in bread wheat, indicating that improving the balance between numerical and mass-related components of productivity may represent a more effective pathway to genetic gain than relying solely on morphological modification. These insights can inform future breeding research and generate hypotheses to be tested for yield stability in multi-environment evaluations. The results revealed that wheat yield formation depends mainly on the synergistic balance between grain number and grain mass rather than on architectural morphology. These findings, within the limits of a single-season, single-location evaluation, suggest that selection emphasis may be placed on genotypes combining higher GPS with stable TWK, but broader inference requires multi-environment validation. This study provided one of the first SEM-based insights into yield architecture in Azerbaijan's wheat germplasm. Because the experiment was conducted in one environment (single year, single location) with one replication, the results should be interpreted as exploratory and context-specific rather than broadly generalisable. A promising direction for future research is the evaluation of the stability of the identified structural relationships between yield components in multi-location, multi-factor trials, taking into account genotype and environment interactions.

## ACKNOWLEDGEMENTS

The authors express sincere gratitude to the staff and administration of the Azerbaijan State Genetic Resources Institute, as well as to all participants and institutional representatives whose support and cooperation contributed significantly to the successful completion of this research.

## FUNDING

None.

## CONFLICT OF INTEREST

None.

## REFERENCES

- [1] Arshad, S., Jamil, M., Ali, A., Saleem, Y., Khan, U.A., Maqsood, M.F., Zulfikar, U., Ali, M.F., Binobead, M.A., & AlMunqedhi, B.M. (2025). Exploring multivariate relationships among seed morphometric and yield-related traits in bread wheat (*Triticum aestivum* L.). *Polish Journal of Environmental Studies*, 34(3), 2003-2010. doi: 10.15244/pjoes/187779.

- [2] Barman, M., Choudhary, V.K., Singh, S.K., Parveen, R., & Gowda, A.K. (2020). Correlation and path coefficient analysis in bread wheat (*Triticum aestivum* L.) genotypes for morpho-physiological traits along with grain Fe and Zn content. *Current Journal of Applied Science and Technology*, 39(36), 130-140. doi: [10.9734/cjast/2020/v39i3631081](https://doi.org/10.9734/cjast/2020/v39i3631081).
- [3] Baye, A., Berihun, B., Bantayehu, M., & Derebe, B. (2020). Genotypic and phenotypic correlation and path coefficient analysis for yield and yield-related traits in advanced bread wheat (*Triticum aestivum* L.) lines. *Soil & Crop Sciences*, 6(1), article number 1752603. doi: [10.1080/23311932.2020.1752603](https://doi.org/10.1080/23311932.2020.1752603).
- [4] Borkhatariya, T.H., Pansuriya, A.G., Godhani, C.M., Mehta, D.R., Vekaria, D.M., Kulkarni, G.U., Madaria, R.B., & Patel, J.B. (2025). Exploring trait relationships: Correlation and path analysis in bread wheat (*Triticum aestivum* L.). *Annual Research & Review in Biology*, 40(4), 58-66. doi: [10.9734/arrb/2025/v40i42221](https://doi.org/10.9734/arrb/2025/v40i42221).
- [5] Chauhan, V., Vimal, S.C., Kumar, P., Kumar, S., & Jha, N. (2025). Correlation and path coefficient analysis for yield and yield-related traits in bread wheat (*Triticum aestivum* L.). *Journal of Advances in Biology & Biotechnology*, 28(7), 1097-1102. doi: [10.9734/jabb/2025/v28i72626](https://doi.org/10.9734/jabb/2025/v28i72626).
- [6] Convention on Biological Diversity. (1992, May). Retrieved from <https://www.cbd.int/convention>.
- [7] Ezici, A.A., Hizli, H., Yaktubay, S., Ay, H., & Oluk, C.A. (2022). Interpretation of morphological and quality characteristics affecting yield in some bread wheat (*Triticum aestivum* L.) genotypes by path, correlation analysis and genetic variability. *SSRG International Journal of Agriculture and Environmental Science*, 9(2), 15-21. doi: [10.14445/23942568/IJAES-V9I2P102](https://doi.org/10.14445/23942568/IJAES-V9I2P102).
- [8] Fang, H., Liu, F., Gu, X., Chen, P., Li, Y., & Li, Y. (2022). The effect of source-sink on yield and water use of winter wheat under ridge-furrow with film mulching and nitrogen fertilisation. *Agricultural Water Management*, 267, article number 107616. doi: [10.1016/j.agwat.2022.107616](https://doi.org/10.1016/j.agwat.2022.107616).
- [9] Fischer, T., Gonzalez, F.G., & Miralles, D.J. (2024). Breeding for increased grains/m<sup>2</sup> in wheat crops through targeting critical period duration: A review. *Field Crops Research*, 316, article number 109497. doi: [10.1016/j.fcr.2024.109497](https://doi.org/10.1016/j.fcr.2024.109497).
- [10] Flavell, R.B. (2023). A framework for improving wheat spike development and yield based on the master regulatory TOR and SnRK gene systems. *Journal of Experimental Botany*, 74(3), 755-768. doi: [10.1093/jxb/erac469](https://doi.org/10.1093/jxb/erac469).
- [11] Grace, J.B. (2009). *Structural equation modeling and natural systems*. Cambridge: Cambridge University Press. doi: [10.1017/CBO9780511617799](https://doi.org/10.1017/CBO9780511617799).
- [12] Havryliuk, I., & Kovalyshyna, H. (2025). Analysis of soft winter wheat hybrids for main morphological and productive traits of the ear. *Scientific Reports of the National University of Life and Environmental Sciences of Ukraine*, 21(2), 9-23. doi: [10.31548/dopovidi/2.2025.09](https://doi.org/10.31548/dopovidi/2.2025.09).
- [13] Javed, M., Ali, A., Kashif, M., Ali, M., Ullah, S., & Alam, A. (2024). Estimation of heritability, genotypic variability and correlations analysis for yield and yield attributing traits among bread wheat (*Triticum aestivum* L.) genotypes. *Journal of Applied Life Sciences and Environment*, 57(1(197)), 91-106. doi: [10.46909/alse-571125](https://doi.org/10.46909/alse-571125).
- [14] Kour, G., Shekhawat, U.S., Mukhram, Komalpreet K., & Appy. (2025). Assessment of genetic variability, correlation and path analysis in wheat (*Triticum aestivum* L.). *Journal of Advances in Biology & Biotechnology*, 28(11), 360-371. doi: [10.9734/jabb/2025/v28i113241](https://doi.org/10.9734/jabb/2025/v28i113241).
- [15] Li, C.C. (1975). *Path analysis: A primer*. Pacific Grove: Boxwood Press.
- [16] Mahpara, S., Imran, M., Khalid, M.A., Fatima, N., Ammar, A., Liaqat, S., & Kanwal, S. (2024). [Correlation and path analysis in different wheat genotypes for grain yield and its related traits: Wheat genotypes for grain yield](https://doi.org/10.31548/dopovidi/2.2025.09). *Pakistan Journal of Scientific and Industrial Research Series B: Biological Sciences*, 67(3), 226-231.
- [17] Milkessa, T.O. (2022). Correlation and path coefficient analysis of traits in bread wheat (*Triticum aestivum* L.) genotypes under drought stress conditions. *American Journal of Plant Biology*, 7(3), 120-126. doi: [10.11648/j.ajpb.20220703.11](https://doi.org/10.11648/j.ajpb.20220703.11).

- [18] Ochagavía, H., Prieto, P., Savin, R., Griffiths, S., & Slafer, G.A. (2018). Earliness per se effects on developmental traits in hexaploid wheat grown under field conditions. *European Journal of Agronomy*, 99, 214-223. doi: [10.1016/j.eja.2018.07.007](https://doi.org/10.1016/j.eja.2018.07.007).
- [19] Rajput, A.G., Misal, A.M., Jadhav, A.V., Shinde, S.A., & Rathod, A.H. (2025). Correlation and path coefficient analysis in bread wheat (*Triticum aestivum* L.). *International Journal of Agriculture and Food Science*, 7(7(F)), 437-440. doi: [10.33545/2664844X.2025.v7.i7f.541](https://doi.org/10.33545/2664844X.2025.v7.i7f.541).
- [20] Sharadhi, G.P., Avinashe, H., Dubey, N., Danalakoti, K., Ghosh, S., Sachan, S., & Choudhary, S. (2023). Correlation and path analyses in F4 segregating populations of bread wheat for grain yield and its attributes. *SABRAO Journal of Breeding and Genetics*, 55(5), 1666-1677. doi: [10.54910/sabrao2023.55.5.19](https://doi.org/10.54910/sabrao2023.55.5.19).
- [21] Slafer, G.A., Savin, R., & Sadras, V.O. (2023). Wheat yield is not causally related to the duration of the growing season. *European Journal of Agronomy*, 148, article number 126885. doi: [10.1016/j.eja.2023.126885](https://doi.org/10.1016/j.eja.2023.126885).
- [22] Ullah, M.I., et al. (2021). Grain yield and correlated traits of bread wheat lines: Implications for yield improvement. *Saudi Journal of Biological Sciences*, 28(10), 5714-5719. doi: [10.1016/j.sjbs.2021.06.006](https://doi.org/10.1016/j.sjbs.2021.06.006).
- [23] Vicentin, L., Canales, J., & Calderini, D.F. (2024). The trade-off between grain weight and grain number in wheat is explained by the overlapping of the key phases determining these major yield components. *Frontiers in Plant Science*, 15, article number 1380429. doi: [10.3389/fpls.2024.1380429](https://doi.org/10.3389/fpls.2024.1380429).
- [24] Wang, K., Shi, L., Zheng, B., & He, Y. (2023). Responses of wheat kernel weight to diverse allelic combinations under projected climate change conditions. *Frontiers in Plant Science*, 14, article number 1138966. doi: [10.3389/fpls.2023.1138966](https://doi.org/10.3389/fpls.2023.1138966).
- [25] Wright, S. (1921). [Correlation and causation](#). *Journal of Agricultural Research*, 20, 557-586.
- [26] Wu, X.-L., Liu, M., Li, C.-S., Mchugh, A.D., Li, M., Xiong, T., Liu, Y.-B., & Tang, Y.-L. (2022). Source-sink relations and responses to sink-source manipulations during grain filling in wheat. *Journal of Integrative Agriculture*, 21(6), 1593-1605. doi: [10.1016/s2095-3119\(21\)63640-7](https://doi.org/10.1016/s2095-3119(21)63640-7).
- [27] Yadav, M., Yadav, V., Singh, U.P., Kumar, K., Gautam, U., & Singh, S.N. (2025). Studies on genetic parameters, correlation coefficient, and path analysis for yield and yield-contributing traits in bread wheat (*Triticum aestivum* L.). *International Journal of Research in Agronomy*, 8(1), 532-539. doi: [10.33545/2618060X.2025.v8.i1Sh.2494](https://doi.org/10.33545/2618060X.2025.v8.i1Sh.2494).
- [28] Zewdu, D., Mekonnen, F., & Geleta, N. (2024). Correlation and path analysis for yield and yield component traits on bread wheat (*Triticum aestivum* L.) genotypes. *International Journal of Bio-resource and Stress Management*, 15(1). doi: [10.23910/1.2024.5040](https://doi.org/10.23910/1.2024.5040).
- [29] Zhang, C., Zheng, B., & He, Y. (2022). Improving grain yield via promotion of kernel weight in high yielding winter wheat genotypes. *Biology*, 11(1), article number 42. doi: [10.3390/biology11010042](https://doi.org/10.3390/biology11010042).
- [30] Zhang, M., Gao, Y., Zhang, Y., Fischer, T., Zhao, Z., Zhou, X., Wang, Z., & Wang, E. (2020). The contribution of spike photosynthesis to wheat yield needs to be considered in process-based crop models. *Field Crops Research*, 257, article number 107931. doi: [10.1016/j.fcr.2020.107931](https://doi.org/10.1016/j.fcr.2020.107931).

## Кореляційний та шляховий аналіз морфологічних ознак у колекції м'якої пшениці за допомогою моделювання структурних рівнянь

### Наргіз Еюбова

Аспірант

Інститут генетичних ресурсів Національної академії наук Азербайджану

AZ1106, просп. Азадлиг, 155, м. Баку, Азербайджан

<https://orcid.org/0009-0008-6615-7234>

### Гамлет Садіков

Доктор біологічних наук, доцент

Інститут генетичних ресурсів Національної академії наук Азербайджану

AZ1106, просп. Азадлиг, 155, м. Баку, Азербайджан

<https://orcid.org/0009-0006-9302-6817>

### Севіль Садігова

Кандидат біологічних наук, провідний науковий співробітник

Інститут генетичних ресурсів Національної академії наук Азербайджану

AZ1106, просп. Азадлиг, 155, м. Баку, Азербайджан

<https://orcid.org/0009-0009-4718-9561>

### Нікет Гасім

Асистент-професор

Карабаський університет

AZ2600, м. Ханкенді, Азербайджан

<https://orcid.org/0000-0002-3976-1283>

**Анотація.** Метою цього дослідження було вивчення структурних детермінант урожайності зерна з однієї рослини у пшениці м'якої (*Triticum aestivum* L.) та оцінка внеску морфологічних і продуктивних ознак. У роботі застосовано методи описової статистики, кореляційний аналіз Пірсона, аналіз шляхових коефіцієнтів і моделювання структурними рівняннями (SEM). Досліджено 53 генотипи з колекції генетичних ресурсів Інституту генетичних ресурсів (м. Баку, Азербайджан) за такими показниками: маса зерна з колоса, маса тисячі зерен (TKW), кількість зерен у колосі (GPS), висота рослини (PH) і довжина головного колоса (SL). Аналіз показав, що врожайність зерна з однієї рослини (GWPP) визначається передусім кількісними (GPS,  $r=0,47$ ) і масовими (TKW,  $r=0,29$ ) показниками, тоді як архітектурні ознаки (PH, SL) мають переважно опосередкований вплив. Моделювання SEM засвідчило, що маса зерна з колоса (GWPS) виступає центральною медіаторною змінною, через яку реалізується вплив GPS і TKW на загальну врожайність. Вісь «маса/стік» (TKW → GWPS) характеризується сильнішим прямим ефектом (стандартизований коефіцієнт = 2,18), ніж архітектурна вісь (0,41), тоді як архітектурні ознаки переважно опосередковано підтримують потенціал «стоку». Побудована модель продемонструвала відмінну відповідність даним ( $\chi^2(3) = 1,26$ ,  $p = 0,74$ ; CFI = 1,00; RMSEA = 0,00; SRMR = 0,036), що підтверджує біологічну інтерпретованість виявлених взаємозв'язків між ознаками. Отримані результати мають практичне значення для селекціонерів і генетиків рослин у процесі добору високоврожайних генотипів пшениці в контрольованих польових умовах і можуть бути використані для оптимізації стратегій підвищення продуктивності пшениці м'якої

**Ключові слова:** *Triticum aestivum* L.; урожайність зерна; класична структура; хімічні обробки; динаміка джерела-поглинач