# Mislabeling latent variables in factor analysis applied to football playing style research: A commentary

Spyridon Plakias[1*,A-F]

[1] Department of Physical Education and Sport Science, University of Thessaly, 42100 Trikala, Greece

***Corresponding author:** Spyridon Plakias; University of Thessaly, Argonauton kai Filellinon, Volos 382 21, Greece; email: spyros_plakias@yahoo.gr

**Received:** 2026-01-25
**Accepted:** 2026-03-23
**Published:** 2026-04-01

A – Research concept and design
B – Collection and/or assembly of data
C – Data analysis and interpretation
D – Writing the article
E – Critical revision of the article
F – Final approval of article





## ABSTRACT

The application of exploratory factor analysis and principal component analysis (PCA) has become increasingly prevalent in football analytics, offering a robust framework for identifying and quantifying playing styles. By condensing a large set of performance indicators into factors, these techniques provide valuable insights into tactical patterns. However, a persistent challenge lies in the subjective labeling of these factors, which can lead to ambiguous or misleading interpretations. This commentary highlights the issue of mislabeling latent variables in football playing style research, drawing on examples from recent studies to illustrate how inaccurate labels can distort both theoretical understanding and practical applications. The implications of poor labeling extend beyond academic discourse, potentially affecting coaches and analysts who rely on data-driven insights for tactical planning. To address these challenges, we propose best practice recommendations emphasizing conceptual clarity, standardized terminology, and expert validation. Accurate and consistent factor labeling is essential to ensure the reliability, comparability, and practical relevance of findings in the evolving field of football performance analysis.

**Keywords:** factor naming, game style, soccer, soccer analytics, sports performance research.

**How to cite this article:** Plakias, S. (2026). Mislabeling latent variables in factor analysis applied to football playing style research: A commentary. *Physical Education and Sports: Studies and Research*, *5*(1), 40-46. https://doi.org/10.56003/pessr.v5i1.693

## INTRODUCTION

The increasing use of multivariate statistical techniques, such as exploratory factor analysis and principal component analysis (PCA), has significantly advanced the study of football playing styles (Plakias et al., 2023). These methods enable researchers to reduce a large set of performance indicators into a smaller number of factors (Rojas-Valverde et al., 2020; Schreiber, 2021), which can then be interpreted as distinct tactical patterns or styles of play (Fernandez-Navarro et al., 2016; Pollard et al., 1988). In this context, the components or factors extracted through these techniques are often interpreted as latent tactical constructs representing playing styles. While the terms "factor" and "latent variable" are sometimes used interchangeably in applied research, strictly speaking factors refer to the statistical dimensions derived from the analysis, whereas latent variables represent the underlying theoretical constructs they aim to capture. While these data-driven approaches offer a powerful means of understanding the complex dynamics of football, a persistent challenge lies in the subjectivity of labeling the extracted factors.

As noted in methodological literature, the interpretation and naming of factors is inherently theoretical and inductive, relying on the researcher's conceptual framework and assumptions (Ford et al., 1986; Watkins, 2018; Williams et al., 2010). This interpretative step reflects issues of construct validity, as factor labels are not determined solely by statistical loadings but also by the theoretical framework through which the variables are conceptually interpreted. In football research, this subjectivity has often led to ambiguous or misleading labels, which do not accurately reflect the variables that load on a factor. This issue not only compromises the scientific validity of studies but also risks misleading practitioners, including coaches and performance analysts, who rely on these analyses for tactical planning and decision-making.

This commentary addresses the recurring issue of mislabeling factors in PCA, using examples from football playing style research. It also offers recommendations for more rigorous and transparent naming practices, emphasizing the need for conceptual clarity and alignment between statistical results and tactical constructs.

### Examples of Questionable Factor Labeling in Football Playing Style Research

Several examples have already been reported by Plakias et al. (2023) in their scoping review on football teams' playing styles. Specifically, the authors highlight cases in which researchers assigned labels to extracted components that do not adequately reflect the variables that load on them. For example, in the study by Gómez et al. (2018), two factors were labeled as set pieces and free kicks, although they in fact referred exclusively to attacking set-pieces and attacking free kicks, respectively. Furthermore, Factor 4 was labeled counter-attack, despite the fact that one of its loading variables is lost balls, which are not related to a team's own counter-attacks but rather to the opponent's play. Therefore, it should either have been clarified that the label refers to the game as a whole (and not to the style of a single team) or, if the intention was to describe only one team's style, a more appropriate term such as transitions should have been used, as it encompasses both offensive and defensive counter-attacking situations. Similarly, in the study by Lago-Peñas et al. (2017), Factors 4 and 5 were both given the general label transitional play, even

though their variables correspond to different phases of the game: Factor 4 is associated with defensive transitions (as indicated by the presence of the variable lost balls), whereas Factor 5 pertains to offensive transitions (as indicated by the presence of the variable picking up free balls). Finally, Ruan et al. (2022) labeled Factors 6 and 8 as defense of goalkeepers, despite the fact that they are not the same. As the authors themselves reported, the two factors yielded different values across various teams.

Recently, in the article by Peng et al. (2025) entitled "Evolution of playing styles in UEFA European championships: trends from 2012 to 2024," there are two points of concern regarding the labeling of factors. Specifically, Factor 1, which includes the variables Total shots, Key passes, Shots on target, Shots from the penalty area, and Shots from outside of the box, is labeled as "Direct Attacking." However, none of these variables inherently indicates a direct style of play. They primarily reflect a team's capacity to generate final attempts, regardless of whether the team employs a direct or possession-based approach. Similarly, Factor 3 is labeled "Wide Play," but it includes Passes at final third, a variable that is not exclusive to wing play and can equally occur through central channels. This broad variable does not, by itself, suggest width-oriented tactics.

All these situations, summarized in Table 1, highlight the importance of ensuring that factor labels are conceptually aligned with both the statistical structure of the data and the tactical realities they aim to represent. Without this alignment, the resulting interpretations risk being misleading for both the scientific community and applied practice.

**Table 1.** Summary of examples of questionable factor labeling in football playing style research

| Study | Factor | Variables | Label | Issue | Suggested Alternative |
|---|---|---|---|---|---|
| Lago-Peñas et al. (2017) | 4 | Lost balls, Lost balls in opponents half | Transitional play | Only a defensive transition can result from losing the ball | Defensive transition |
| Lago-Peñas et al. (2017) | 5 | Picking up free balls, Picking up free balls in opponents half | Transitional play | Only an attacking transition can result from picking up free balls | Attacking transition |
| Gómez et al. (2018) | 4 | Ball possession 5-15s, Lost balls, Recovered balls, Picking free ball, Counter attacks | Counter attack | Lost balls can lead to a counterattack by the opposing team but not by the reference team | Transition play |
| Gómez et al. (2018) | 5 | Free-kick crosses, Set | Set piece | They all concern set pieces performed by | Attacking set pieces |

**Table 1.** (Continued)

| | | | | | |
|---|---|---|---|---|---|
| | | pieces attacks, Set piece attacks with shot | | the reference team (not the opponents) and therefore the name can be made more specific | |
| Gómez et al. (2018) | 8 | Free-kick shots, Free-kick shots on target | Free-kick | They all concern free-kicks performed by the reference team (not the opponents) and therefore the name can be made more specific | Attacking free-kicks |
| Ruan et al. (2022) | 6 | Goalkeeper smother | Defense of goalkeeper | Same label was assigned to the factors even though they are loaded by different variables and, consequently, have different values | Reactive goalkeeper defense (smothering) |
| Ruan et al. (2022) | 8 | Goalkeeper claim | Defense of goalkeeper | | Proactive goalkeeper control (ball claiming) |
| Peng et al. (2025) | 1 | Total shots, Key passes, Shots on target, Shot from penalty area, Shot from outside of box | Direct attacking | The variables mainly reflect attacking output (e.g., shots and key passes) rather than a direct attacking style, as no indicators typically associated with direct play (such as long passes, forward passes, short pass sequences, high-speed transitions, or counterattacks) are included | Offensive production or Shot creation |
| Peng et al. (2025) | 3 | Crosses, Corners, Passes at final third | Wide play | The variables do not specifically represent wide attacking play. While crosses are typically associated with wide areas, corners may result from actions in central zones (e.g., blocked shots), and passes in the final third reflect general attacking activity rather than wide play | Final third attacking activity |

### Implications of Poor Factor Labeling

Incorrect or ambiguous labeling of factors in football playing style research can have multiple adverse consequences. From a scientific perspective, misleading labels distort the theoretical understanding of tactical situations, reducing the validity and reliability of the findings. Researchers may also draw incorrect inferences about a team's style of play, which can hinder the comparability of results across studies. Furthermore, poor labeling undermines the reproducibility and transparency of research, as future studies may adopt or replicate inaccurate terminology, perpetuating conceptual confusion rather than contributing to a unified framework of football tactics. For practitioners, such as coaches and performance analysts, misinterpreted factor labels can result in misguided tactical decisions, inappropriate training focus, or flawed match preparation.

### Recommendations for Best Practice

To minimize the risks associated with poor factor labeling, researchers should adopt a set of rigorous and transparent practices. First, the interpretation of factors should be firmly grounded in both the statistical structure of the data (e.g., high-loading variables, factor correlations) and the theoretical framework of football tactics. Furthermore, researchers are encouraged to align their labeling strategies with standardized terminology where possible, fostering consistency across studies and enabling the development of a shared conceptual framework for playing style analysis. Avoiding overly broad or ambiguous labels is crucial; terms should reflect the common tactical meaning of the variables rather than arbitrary or misleading descriptors. Finally, achieving all the above will be facilitated by cross-validation of factor labels through expert panels (including coaches, performance analysts, and sport scientists), which can help ensure that the chosen terms accurately represent the underlying constructs.

In addition, transparency in reporting is essential. Researchers are encouraged to provide full factor loading matrices (e.g., in supplementary materials) to allow readers to evaluate the basis of factor interpretation. Furthermore, future studies could employ confirmatory factor analysis to examine whether the proposed interpretations of playing style factors remain consistent across different datasets or competitions.

## CONCLUSIONS

The accurate labeling of factors derived from PCA is crucial for advancing the scientific understanding of football playing styles. Mislabeling, as evidenced by several recent studies, not only undermines the interpretability and validity of research findings but also risks creating conceptual confusion that may mislead both scholars and practitioners. Establishing clear, theoretically grounded, and data-informed naming practices is therefore essential for ensuring that the identified factors genuinely represent the tactical constructs they are intended to capture. By adopting rigorous methodological standards, aligning terminology with established tactical frameworks, and validating labels through expert consultation, researchers can enhance both the reliability and practical utility of PCA-based factor extraction in football analytics. Future work should prioritize the development of consensus

guidelines for naming latent variables, thereby fostering consistency, comparability, and meaningful knowledge advancement in the field of playing style research.

## AKCNOWLEDGMENT

The author would like to thank and congratulate the researchers and journals whose work is reviewed and discussed in this manuscript.

## AI DISCLOSURE STATEMENT

The author acknowledges the use of ChatGPT (OpenAI) solely for improving the clarity and quality of the English language.

## DATA AVAILABILITY

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

## FUNDING

No funding was received for the preparation of this manuscript.

## CONFLICT OF INTEREST

No potential conflict of interest was reported by the author.

## PUBLISHER'S NOTE

All claims expressed in this article are solely those of the author and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or a claim its manufacturer may make, is not guaranteed or endorsed by the publisher.

## REFERENCES

Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P. R., & McRobert, A. P. (2016). Attacking and defensive styles of play in football: analysis of Spanish and English elite teams. *Journal of sports sciences, 34*(24), 2195-2204. https://doi.org/10.1080/02640414.2016.1169309

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*(2), 291-314. https://doi.org/10.1111/j.1744-6570.1986.tb00583.x

Gómez, M.-Á., Mitrotasios, M., Armatas, V., & Lago-Peñas, C. (2018). Analysis of playing styles according to team quality and match location in Greek professional football. *International Journal of Performance Analysis in Sport, 18*(6), 986-997. https://doi.org/10.1080/24748668.2018.1539382

Lago-Peñas, C., Gómez-Ruano, M., & Yang, G. (2017). Styles of play in professional football: an approach of the Chinese Football Super League. *International*

*Journal of Performance Analysis in Sport, 17*(6), 1073-1084. https://doi.org/10.1080/24748668.2018.1431857

Peng, W., Ribeiro, J., Wang, M., & Gómez, M. Á. (2026). Evolution of playing styles in UEFA European championships: trends from 2012 to 2024. *International Journal of Performance Analysis in Sport,* 1-16. https://doi.org/10.1080/24748668.2025.2536954

Plakias, S., Moustakidis, S., Kokkotis, C., Tsatalas, T., Papalexi, M., Plakias, D., Giakas, G., & Tsaopoulos, D. (2023). Identifying football teams' styles of play: a scoping and critical review. *Journal of Functional Morphology and Kinesiology, 8*(2), 39. https://doi.org/10.3390/jfmk8020039

Pollard, R., Reep, C., & Hartley, S. (2013). The quantitative comparison of playing styles in soccer. In *Science and Football (Routledge Revivals)* (pp. 309-315). Routledge. https://doi.org/10.4324/9780203720035

Rojas-Valverde, D., Pino-Ortega, J., Gómez-Carmona, C. D., & Rico-González, M. (2020). A systematic review of methods and criteria standard proposal for the use of principal component analysis in team's sports science. *International Journal of Environmental Research and Public Health, 17*(23), 8712. https://doi.org/10.3390/ijerph17238712

Ruan, L., Ge, H., Gómez, M. Á., Shen, Y., Gong, B., & Cui, Y. (2023). Analysis of defensive playing styles in the professional Chinese Football Super League. *Science and Medicine in Football*, *7*(3), 279–287. https://doi.org/10.1080/24733938.2022.2099964

Schreiber, J. B. (2021). Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy, 17*(5), 1004-1011. https://doi.org/10.1016/j.sapharm.2020.07.027

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219-246. https://doi.org/10.1177/0095798418771807

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*, 1-13. https://doi.org/10.33151/ajp.8.3.93