International Journal of Performance Analysis of Sport
2009, 9, 8-25.

Developing a New Method for Team Match Performance Analysis in Professional Soccer and Testing its Reliability

A. Tenga, D. Kanstad, L. T. Ronglan and R. Bahr

Norwegian School of Sport Sciences, P. O. Box 4014 Ullevaal Stadion, N-0806 Oslo.

Abstract

The purpose of this study was to develop a new method for analysis of team match performance in soccer and test its reliability from videotapes. Team match performance analysis, which includes qualitative evaluation of different dimensions of match performance involved in the opponent interaction, was developed by using 22 multidimensional categorical variables. This is thought to provide a more valid assessment of team match performance, since it is considered to be a product of the confrontation between two teams. To test inter- and intra-observer reliability, a random sample of 200 team possessions from videos of 163 out of 182 (90%) matches played in the Norwegian professional league in 2004 was analysed. Kappa coefficients (κ) values of 0.81-1.0 were considered very good, 0.61-0.80 good, 0.41-0.60 moderate, 0.21-0.40 fair, and <0.21 poor. Kappa values were considerably better for the intra-observer test (16 (73%) very good, (23%) good and 1 (4%) moderate) than the inter-observer test (7 (32%) very good, 5 (23%) good, 7 (32%) moderate, 2 (9%) fair and 1 (4%) poor of 22 variables). A new method, representing a potentially valuable tool for more valid assessment of team match performance, has been developed as a reliable method for most variables used.

Keywords: Reliability, validity, multidimensional qualitative evaluation, opponent interaction, soccer, match performance analysis

1. Introduction

Match performance analysis is widely used as a method for studying technical, tactical and physical aspects of player and team performance in a soccer match (McGarry & Franks, 2003). Therefore, ensuring validity and adequate reliability of the collected data is important for match performance analysis to meet its intents and purposes effectively. Yet, this scientific basis for match performance analysis has neither been established nor fully pursued in most studies.

According to the analytical framework based on 'dynamical configuration of play', the opposition creates the unexpected in a match, necessitating constant adaptation to constraints caused by the confrontation between two teams (Elias & Dunning, 1966; Grehaigne *et al*., 1997). Hence, match performance analysis must consider the interaction between the two opposing teams to be a more valid analysis. It is therefore surprising that

most studies on match performance do not consider the assessment of the opponent interaction in their analyses (e.g. James *et al*., 2002; Hughes & Churchill, 2004; Jones *et al*., 2004; Taylor *et al*., 2004, 2005; Konstadinidou & Tsigilis, 2005; Tucker *et al*., 2005; Hughes & Snook, 2006). Indeed, these studies often use uni-dimensional quantitative data (frequency or counts of match play events), which are incapable of describing different dimensions of match performance taking place during opponent interaction. In support, Olsen and Larsen (1997) wrote that it is not easy to quantify all analysis variables of match performance in soccer, especially the important ones.

The opponent interaction involves multiple dimensions of match performance. These include temporal and spatial dimensions, whose direct measurements are difficult to obtain. Since categorical data based on multidimensional qualitative evaluation permits the inclusion of data from the qualitative evaluation of different dimensions of match performance, it's use, instead of uni-dimensional frequency data, may improve our ability to describe a soccer match play action (Grehaigne *et al.*, 2001; Hughes & Bartlett, 2002; Suzuki & Nishijima, 2004). For example, a variable such as "team possession type" could be used in an attempt to describe the two opposing offensive tactics, namely *counter attack* ("direct play") and *elaborate attack* ("possession play") by using categories indicating the degree of offensive directness. Similarly, a variable "defensive pressure", could be used to describe degrees of *loose pressure* to *tight pressure*, possibly through ordered categories of estimated pressing distances. Likewise, ordered categories indicating the number of touches per ball involvement could be used to describe degrees of high tempo to low tempo of play. Hence, multidimensional qualitative evaluation may be achieved by converting frequency data of different factors of match performance, well recognised in practice, into ordered and non-ordered categorical data.

It is possible to assess the opponent interaction when using a *match play situation* as the basic unit of analysis rather than *a team* (or *a player*), and therefore a *team possession* was used as the unit of analysis in this study. Since match play situations emerge from the interplay of *play* and *counter play* produced by both teams (Grehaigne & Godbout, 1995; Grehaigne *et al*., 1999), it enables breaking down of a match play action without losing its confrontational nature (e.g. Harris & Reilly, 1988; Grehaigne, 1991; Suzuki & Nishijima, 2004; Seabra & Dantas, 2006). Moreover, the dynamic analysis of match performance could be achieved by a continuous analysis, offensively and defensively, of each attempt to win or receive the ball in all ball involvements from the start to the end of a team possession. The frequencies of each category within a team possession could then be summed up and used to characterize each team possession according to offensive and defensive variables.

Regarding the reliability of the collected data, the fact that data variability limits the ability to detect a true difference between performances (Atkinson & Nevill, 1998; Hughes *et al*., 2004; O'Donoghue, 2007) makes adequate reliability vital for match performance analysis to be a valid method. Hughes *et al.* (2004) maintain that it is very important to test any new analysis system and ensure adequate reliability (objectivity and accuracy) of the collected data. However, this is not the case in most match performance studies. For example, all 15 studies on match performance analysis in soccer that report reliability results found in the literature fail to account for how they estimated sample size for their reliability studies (e.g.

Hughes *et al.*, 1998; Luhtanen *et al.*, 2001; Scolding *et al.*, 2004; Suzuki & Nishijima, 2004; Bloomfield *et al.*, 2005; Hughes & Reed, 2005; Seabra & Dantas, 2006). This is especially important since some events that make typical analysis variables naturally occur either more or less frequently than others during a match play action and, therefore, different analysis variables often demand different sizes to make an adequate sample (Cooper *et al.*, 2007). In addition, these studies often provide inadequate and unclear information concerning the reliability testing procedures (Hughes *et al.*, 2004). The fact that all these studies include reliability tests only as a part of the main study may also contribute to the lack of important details evident in the reported reliability.

Thus, the aim of this study was to develop a new method for team match performance analysis in soccer which includes an assessment of opponent interaction and multidimensional categorical data and to test its reliability from videotapes.

2. Methods

2.1 Material

Videotapes from 163 out of 182 (90%) matches played in the Norwegian professional league in 2004 were used. This video material was obtained from the Norwegian Broadcasting Corporation (NRK) in digital betacam video format with a conventional TV production. The Norwegian professional league involves 14 teams and follows a double round robin competition format, which means that each team played 26 matches, 13 home and 13 away. To obtain a random sample of 200 team possessions we first selected ten matches from the 163 available on video by choosing every 16^{th} match on the original list. Each match was then assigned a computer-generated random decimal number between 0 and 1, which was multiplied by 86 to indicate the beginning (in minutes) of a match period from which a total of 20 consecutive team possessions would be extracted. This was based on the assumption that 20 consecutive team possessions lasts 6.5 minutes on average, and that there is 2-3 minutes of extra time added to every match.

2.2. Design

A soccer coach and researcher (AT) experienced in match performance analysis and a soccer coach and masters student (DK) performed the analyses. The student underwent a four-week intensive training period in match performance analysis during pilot testing. The two observers then independently analysed the sample to test inter-observer reliability, and one of them (AT) repeated the analysis after three weeks to test intra-observer reliability.

2.3. Team match performance analysis

A team possession was used as the basic unit of analysis and was defined according to Pollard and Reep (1997, p. 542):

> *"A team possession starts when a player gains possession of the ball by any means other than from a player of the same team. The player must have enough control over the ball to be able to have a deliberate influence on its subsequent direction. The team possession may continue with a series of passes between players of the same team but ends immediately when one of the following events*

occurs: a) the ball goes out of play; b) the ball touches a player of the opposing team (e.g. by means of a tackle, an intercepted pass or a shot being saved). A momentary touch that does not significantly change the direction of the ball is excluded; c) an infringement of the rules takes place (e.g. a player is offside or a foul is committed).”

Team match performance analysis allows match performance to be described using 22 categorical variables, each with three to seven ordered and non-ordered categories (Table 1, 2, 3 & 4). These were selected based on the “principles of play” explaining players’ functional roles (Worthington, 1980) and the structure of the game (Olsen *et al.*, 1994) as a relationship between space and time in soccer. This methodology characterizes each team possession based on playing tactics; both offensive tactics (18 variables), including main and secondary tactics (Table 1 & 2); and defensive tactics (three zone-oriented defensive variables), including defensive pressure, defensive backup and defensive cover (Table 3); and team possession outcomes (one variable), including categories of dichotomy and discrete outcomes (Table 4).

The nature of analysis in 15 out of 22 variables was qualitative evaluation, e.g. team possession type, space utilization, and defensive cover. The other five variables were analysed by qualitative evaluation assisted by natural lines on the pitch (e.g. starting zone) and two variables were analysed by simply counting on-the-ball events (i.e. pass number and playing tempo). The analysed data were saved in SPSS (version 15.0, SPSS Inc., Chicago). The study was approved by the Norwegian Social Science Data Services (NSD).

2.4. Sample size

The number of team possessions (n) needed for this reliability study was calculated by using the formula for Kappa measure of agreement (κ): $(SE\ (\kappa))^2 = p_o(1 - p_o) / n(1 - p_e)^2$. The observed proportional agreement (p_o) and expected proportional agreement (p_e) were estimated by using data from a pilot study, where the intra-observer reliability for each variable was tested on 30 randomly selected team possessions. A limit of variability (SE (κ)) was set to ≤ 0.05, meaning that the confidence interval (CI) for correlation coefficient values (κ) could be interpreted as $\kappa \pm 1.96 * 0.05$. This implies that there is a 95% probability that the κ value for a particular variable lies between the two limits of CI. Based on these estimates, a value of n for each variable was calculated and from them a sample size of 200 team possessions was estimated to ensure a 95% CI for κ for all variables (Altman, 1991).

Table 1. *Descriptions and category definitions for the main offensive variable and two related variables used in the team match performance analysis.*

Variables and categories
1. Team possession type (four categories, two ordered)
Def. Degree of offensive directness by levels of utilization or creation of imbalance in the opponent's defence to achieve penetration (i.e. how quick penetration is attempted after ball winning). Penetration is achieved when a pass goes towards the opponent's goal past opponent player(s) while maintaining high degree of control over the ball. High degree of control over the ball means enough space and time that makes it easier to perform intended actions on the ball.
A. Counter attack ("direct play"): starts by winning the ball in play and progresses by either a) utilizing or attempting to utilize a degree of imbalance from start to the end, or b) creating or attempting to create a degree of imbalance from start to the end by using early (i.e. 1st or 2nd, evaluated qualitatively) penetrative pass or dribble. Utilizing degree of imbalance means seeking penetration in such a way that a defending team fails to regain high degree of balance from start to the end of team possession. Counter attacks progress relatively fast.
B. Set play: starts by a set play and finishes while players still are more in original set play grouping. In case team possession takes longer time and finishes while players' positions are no longer influenced by original set play grouping, a set play becomes elaborate attack with a set play-start. Set plays often take relatively short time.
C. Elaborate attack ("possession play"): starts by either winning the ball in play or a set play and progresses either a) without utilizing or attempting to utilize a degree of imbalance, or b) by creating or attempting to create a degree of imbalance by using late (3rd or later, evaluated qualitatively) penetrative pass or dribble. Not utilizing a degree of imbalance means seeking penetration in such a way that a defending team manages to regain high degree of balance before the end of team possession. Elaborate attacks often progress relatively slow.
D. Other: team possession that fails to be registered as counter attack or elaborate attack or set play. In addition, team possession that starts by winning the ball in play, but (i) finishes too fast to show a clear attempt to seek penetration or (ii) with no intention to seek penetration, for example during ball clearances, time-wasting tactics and fair play gesture or (iii) shows no entire action due to filming error.
2. Elaborate attack start-type (four non-ordered categories)
Def. Type of starting elaborate attack team possession.
A. Counter attack-start: elaborate attack team possession starts by winning the ball in play.
B. Set play-start: elaborate attack team possession starts by a set play.
C. Not applicable: team possession registered as counter attack, set play, or other.
D. Other
3. Set play start-type (four non-ordered categories)
Def. Quickness of starting set play team possession.
A. Delayed: delay start that allows a defending team to have enough time to establish a balanced defence.
B. Fast: fast start that denies a defending team enough time to establish a balanced defence.
C. Not applicable: team possession starts by winning the ball in play.
D. Other

Table 2. *Descriptions and category definitions for 15 secondary offensive variables used in the team match performance analysis.*

Variables and categories
1. Starting zone (six categories, five ordered)
Def. Area across the playing field in which team possession starts (Figure 1).
A. First third: 1/3 of the playing field estimated from own goal line to middle third 1.
B. Middle third 1: first half of the middle third area estimated from end of the first third to midline.
C. Middle third 2: second half of the middle third area estimated from midline to final third.
D. Final third: 1/3 of the playing field estimated from end of the middle third 2 to opponent's goal line, excluding score box.
E. Score box: Area in front of the opponent's goal defined as an imaginary prolongation of the penalty area from 16 m to 30 m line estimated distance from opponent's goal line.
F. Other
2. Starting corridor (five categories, four ordered)
Def. Area along the playing field in which team possession starts (Figure 1).
A. Right: Area from imaginary line joining right sides of the penalty areas when facing the opponent's goal to right sideline.
B. Central right: Area from imaginary midline along the field to imaginary line joining right sides of the penalty areas when facing the opponent's goal.
C. Central left: Area from imaginary line joining left sides of the penalty areas when facing the opponent's goal to imaginary midline along the field.
D. Left: Area from left sideline to imaginary line joining left sides of the penalty areas when facing the opponent's goal.
E. Other

Table 2 *(continued).*

Variables and categories
3. Starting climate (five categories, two ordered) Def. Opponent's degree of control over the ball prior to losing possession at the end of a preceded team possession. A. High opponent control: starts by winning the ball in play following the opponent's high degree of control over the ball in play. High degree of control over the ball means enough space and time that makes it easier to perform intended action on the ball. B. SP against: starts by winning the ball in play following a set play team possession by the opposing team. C. SP for: starts by a set play after the ball has been out of play. D. Low opponent control: starts by winning the ball in play following the opponent's low degree of control over the ball in play. Low degree of control over the ball means actions with higher risk to lose the ball or not enough space and time that makes it more difficult to perform intended action on the ball. E. Other
4. Player mobility (seven categories, five ordered) Def. Forward runs i.e. running towards the opponent's goal prior to the moment of winning or receiving the ball and otherwise for non-forward runs. A. Forward: only forward runs. B. More forward: greater number of forward than non-forward runs. C. Neutral mobility: equal number of forward and non-forward runs. D. More non-forward: greater number of non-forward than forward runs. E. Non-forward: only non-forward runs. F. Not applicable: Set play team possession without ball reception. G. Other
5. Pass number (six categories, four ordered) Def. Series of passes between players of the attacking team. A. Very low: 1 or 2 passes per team possession. B. Low: 3 passes per team possession. C. High: 4 passes per team possession. D. Very high: 5 or more passes per team possession. E. Not applicable: team possession without a pass. F. Other
6. Playing tempo (six categories, five ordered) Def. Number of touches per ball involvement including set play starting and ball winning at the beginning of team possession. A. High: 1 or 2 touches. B. More high: greater number of high than low tempo involvements. C. Neutral tempo: equal number of low and high tempo involvements. D. More low: greater number of low than high tempo involvements. E. Low: 3 or more touches. F. Other
7. Pass length (seven categories, five ordered) Def. Long passes i.e. 30 m or more estimated distance and shorter estimated distances for short passes. A. Long: only long pass. B. More long: greater number of long than short passes. C. Neutral pass length: equal number of long and short passes. D. More short: greater number of short than long passes. E. Short: only short pass. F. Not applicable: team possession without a pass. G. Other
8. Pass penetration (seven categories, five ordered) Def. Penetrative passes i.e. passes towards the opponent's goal past opponent player(s) while maintaining control over the ball and otherwise for non-penetrative passes. A. Penetrative: only penetrative pass. B. More penetrative: greater number of penetrative than non-penetrative passes. C. Neutral pass penetration: equal number of penetrative and non-penetrative passes. D. More non-penetrative: greater number of non-penetrative than penetrative passes. E. Non-penetrative: only non-penetrative pass. F. Not applicable: team possession without a pass. G. Other

Table 2 *(continued).*

Variables and categories

9. Dribble penetration (seven categories, five ordered)
Def. Penetrative dribbles (including run with the ball), i.e. dribbles towards the opponent's goal past opponent player(s) while maintaining control over the ball and otherwise for non-penetrative dribbles.
A. Penetrative: only penetrative dribble.
B. More penetrative: greater number of penetrative than non-penetrative dribbles.
C. Neutral dribble penetration: equal number of penetrative and non-penetrative dribbles.
D. More non-penetrative: greater number of non-penetrative than penetrative dribbles.
E. Non-penetrative: only non-penetrative dribble.
F. Not applicable: team possession without dribbling.
G. Other

10. Skill level (seven categories, five ordered)
Def. Advanced skills i.e. successful dribble (including penetrative run with the ball), wall pass, and overlap) and non-advanced skills i.e. ordinary pass and unsuccessful advanced skill.
A. Advanced: only successful advanced skill.
B. More advanced: greater number of successful advanced than non-advanced skills.
C. Neutral skill level: equal number of successful advanced and non-advanced skills.
D. More non-advanced: greater number of non-advanced than successful advanced skills.
E. Non-advanced: only non-advanced skill.
F. Not applicable: team possession without a pass or dribbling.
G. Other

11. Space utilization (seven categories, five ordered)
Def. Space passes i.e. passes towards a space further than receiver's immediate reach and foot passes i.e. passes towards a player, evaluated from the moment of making the pass.
A. Space pass: only space pass.
B. More space pass: greater number of space- than foot passes.
C. Neutral utilization: equal number of space- and foot passes.
D. More foot pass: greater number of foot- than space passes.
E. Foot pass: only foot pass.
F. Not applicable: team possession without a pass.
G. Other

12. Ball possessing (seven categories, five ordered)
Def. Forward possessing i.e. player faces towards opponent's goal either prior to or immediately after his 1st touch at the moment of winning or receiving the ball and otherwise for non-forward possessing.
A. Forward: only forward possessing.
B. More forward: greater number of forward than non-forward possessing.
C. Neutral possessing: equal number forward than non-forward possessing.
D. More non-forward: greater number of non-forward than forward possessing.
E. Non-forward: only non-forward possessing.
F. Not applicable: Set play team possession without ball reception.
G. Other

13. Centre pass (six non-ordered categories)
Def. Pass intended for goal-assist made towards score box from the side corridor area of the playing field (Figure 1).
A. Late: goal-assist from the side corridors' area between prolongation of 16 m line and opponent's goal line.
B. Early: goal-assist from the side corridors' area before prolongation of 16 m line.
C. Set play: set play from the side corridors.
D. Not applicable: team possession without a centre pass.
E. Multiple: team possession with more than one centre pass.
F. Other

14. Keeper involvement (six non-ordered categories)
Def. Type of keeper involvement.
A. Back pass: keeper's involvement following a back pass from his own players.
B. Save: keeper's involvement following a scoring attempt from opponent players.
C. Goal kick: keeper's involvement following a ball-out-of-play situation.
D. Free kick: keeper's involvement following an infringement of the playing rules.
E. Not applicable: only out field players involved.
F. Other

Table 2 *(continued).*

Variables and categories
15. Regain possession (five categories, three ordered)
Def. Gaining back control over the ball before the opposing team established its possession.
A. Single: 1 regain possession per team possession.
B. Double: 2 regains possession per team possession.
C. Multiple: 3 or more regains possession per team possession.
D. Not applicable: team possession without regain possession.
E. Other

Table 3. *Descriptions and category definitions for three zone-oriented defensive variables used in the team match performance analysis.*

Variables and categories
1. Defensive pressure (seven categories, five ordered)
Def. Distance between a player with the ball (1st attacker) and an immediate pressing opponent player(s) (1st defender(s)), keeper excluded, at each moment of attempting winning or receiving the ball.
A. Loose ("imbalanced"): only when 1st defender is estimated to be more than 1.5 m.
B. More loose (less "imbalanced"): greater number of loose than tight pressure.
C. Neutral pressure: equal number of tight and loose pressure.
D. More tight (less "balanced"): greater number of tight than loose pressure.
E. Tight ("balanced"): only when 1st defender is estimated to be within 1.5 m.
F. Not applicable: when (i) no pressing opponent player (1st defender) or (ii) set play team possession without ball reception.
G. Other
2. Defensive backup (seven categories, five ordered)
Def. Immediate opponent player(s) supporting 1st defender often from behind (2nd defender(s)), keeper excluded, at each moment of attempting winning or receiving the ball except in 'war' zone. 'War' zone means group duel in front of the goal typically following a pass made towards the score box (Figure 1).
A. Absent ("imbalanced"): only without 2nd defender within 5 m estimated distance from 1st defender.
B. More absent (less "imbalanced"): greater number of without than with 2nd defender.
C. Neutral backup: equal number of with and without 2nd defender.
D. More present (less "balanced"): greater number of with than without 2nd defender.
E. Present ("balanced"): only with 2nd defender within 5 m estimated distance from 1st defender.
F. Not applicable: when (i) no 1st defender or (ii) backup in 'war' zone or (iii) set play team possession without ball reception.
G. Other
3. Defensive cover (seven categories, five ordered)
Def. Opponent player(s) guarding space away from the ball often behind 1st defender(s) and/or 2nd defender(s) (3rd defender(s)), keeper excluded, at each moment of attempting winning or receiving the ball.
A. Absent ("imbalanced"): only without 3rd defender(s) behind 1st and/or 2nd defender(s).
B. More absent (less "imbalanced"): greater number of without than with 3rd defender(s).
C. Neutral cover: equal number of with and without 3rd defender(s).
D. More present (less "balanced"): greater number of with than without 3rd defender(s).
E. Present ("balanced"): only with 3rd defender(s) behind 1st and/or 2nd defender(s).
F. Not applicable: when (i) no 1st and 2nd defender or (ii) set play team possession without ball reception.
G. Other

Table 4. *Descriptions and category definitions for the team possession outcomes used in the team match performance analysis.*

Variables and categories
1. Team possession outcome (seven ordered categories)
Def. Degree of offensive success by dichotomy and discrete levels of effectiveness.
I. Score box: Levels of offensive effectiveness within the score box (Figure 1).
A. Goal scoring: scoring attempt ending with a goal approved by a referee.
B. Scoring opportunity: scoring attempt with relatively high scoring probability (e.g. from shorter distances, from wider angles, with poor keeper positioning) as well as with near-scoring situations such as corner kick direct on crossbar.
C. Score box possession: entry into score box with high degree of control over the ball or set play given to the attacking team as a result of entry into score box. High degree of control over the ball means enough space and time that makes it easier to perform intended action on the ball.
II. Not score box: Levels of offensive effectiveness outside the score box (Figure 1).
D. No score box possession: entry into score box with low degree of control over the ball. Low degree of control over the ball means not enough space and time that makes it more difficult to perform intended action on the ball.
E. Final third: ending up in the final third area of the playing field.
F. Middle third: ending up in the middle third area of the playing field.
G. First third: ending up in the first third area of the playing field.

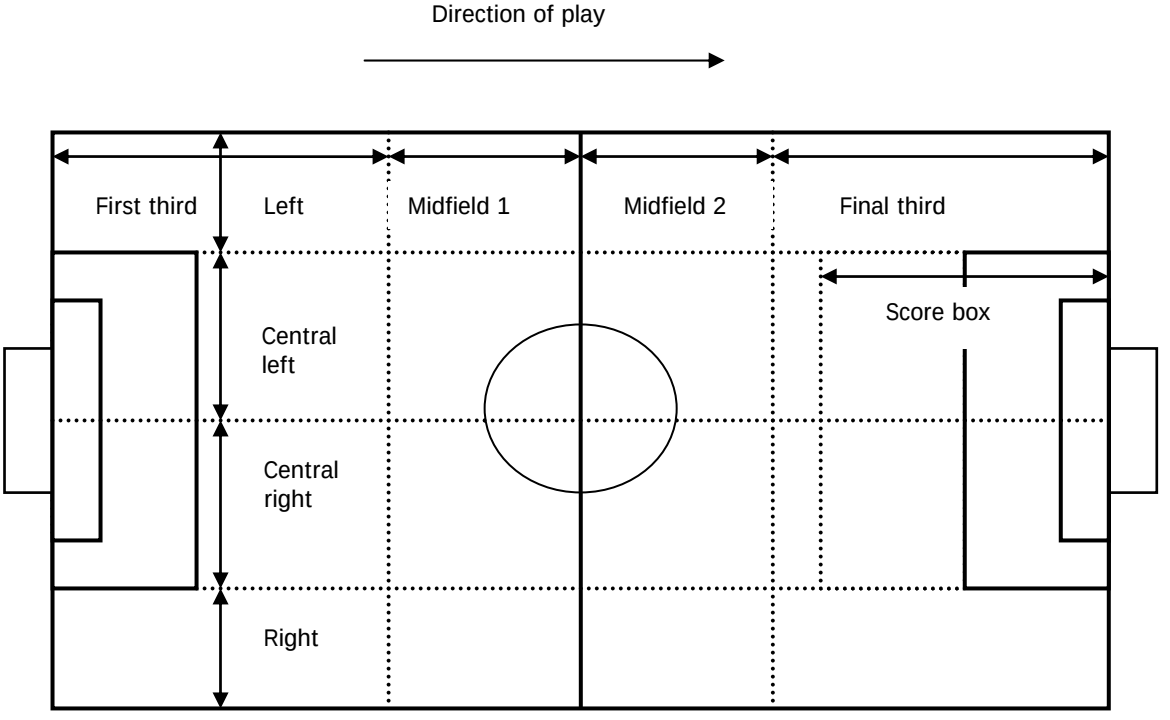

Figure 1. *Zones and corridors of the playing field. Zones include first third, middle third 1, midfield 2, final third, and score box, while corridors include right, central right, central left, and left corridor.*

2.5. Statistical analysis

Kappa correlation coefficients (κ) were calculated for inter-observer and intra-observer reliability. The test was run twice, first for all categories (*all*) and second for only collapsed ordered categories (order*ed*). The test on *all* categories was intended to characterise all match performances within variables, while the test on ordered categories was intended to

simply characterise match performances into two or three ordered categories (degrees of performance). These were obtained after excluding non-ordered categories and collapsing ordered categories without losing practical meaning. κ values of 0.81-1.0 are generally interpreted as very good, 0.61-0.80 as good, 0.41-0.60 as moderate, 0.21-0.40 as fair, and less than 0.21 as poor (Altman, 1991).

3. Results

The κ values were considerably better for the intra-observer than the inter-observer tests when *all* categories were included. For intra-observer tests, κ values were very good for 16 (73%), good for 5 (23%) and moderate for 1 (4%) of 22 variables. For the inter-observer tests, κ values were very good for 7 (32%), good for 5 (23%), moderate for 7 (32%), fair for 2 (9%) and poor for 1 of 22 (4%) variables (Table 5).

Table 5. Kappa correlation coefficients (κ) values for inter- and intra-observer agreement for all categories (*all*) and for only collapsed ordered categories (*ordered)*.

	κ for inter (n)		κ for intra (n)	
Variables and categories	All	*Ordered*	All	*Ordered*
Team possession type	0.59 (191)	0.63 (126)	0.94 (200)	0.92 (160)
Elaborate attack start-type	0.41 (193)	*	0.93 (197)	*
Set play start-type	0.82 (196)	*	0.90 (200)	*
Starting zone	0.85 (200)	0.85 (200)	0.92 (200)	0.94 (200)
Starting corridor	0.85 (200)	0.86 (200)	0.90 (200)	0.93 (200)
Starting climate	0.73 (198)	0.48 (100)	0.92 (200)	0.83 (104)
Player mobility	0.42 (199)	0.44 (182)	0.65 (200)	0.67 (182)
Pass number	0.84 (200)	0.87 (195)	0.94 (200)	0.97 (197)
Playing tempo	0.90 (200)	0.95 (198)	0.92 (199)	0.95 (199)
Pass length	0.79 (200)	0.84 (195)	0.92 (200	0.92 (197)
Pass penetration	0.73 (199)	0.81 (195)	0.84 (200)	0.88 (197)
Dribble penetration	0.57 (197)	0.22 (57)	0.78 (199)	0.74 (60)
Skill level	0.25 (187)	0.26 (191)	0.81 (199)	0.81 (199)
Space utilization	0.49 (199)	0.64 (195)	0.82 (199)	0.87 (196)
Ball possessing	0.54 (200)	0.65 (182)	0.69 (200)	0.77 (182)
Centre pass	0.83 (200)	*	0.94 (200)	*
Keeper involvement	0.88 (199)	*	0.89 (199)	*
Regain possession	0.75 (200)	0.69 (38)	0.88 (200)	0.74 (46)
Defensive pressure	0.57 (200)	0.68 (175)	0.72 (200)	0.82 (174)
Defensive backup	0.19 (199)	0.24 (174)	0.59 (200)	0.57 (173)
Defensive cover	0.35 (200)	0.27 (175)	0.71 (199)	0.67 (173)
Team possession outcome	0.77 (200)	0.75 (200)	0.91 (200)	0.94 (200)

*Kappa values could not be computed because these variables include only non-ordered categories.

The similar distribution of κ values for both intra-observer agreement and inter-observer agreement was registered with improved inter-observer agreement for most variables when only collapsed ordered categories were included. For the intra-observer tests on *ordered* categories, κ values were very good for 12 (67%), good for 5 (28%), moderate for 1 (5%), while for the inter-observer tests κ values were very good for 6 (33%), good for 6 (33%), moderate for 2 (11%) and fair for 4 (22%) of 18 variables (Table 5).

The variables involving qualitative evaluation alone recorded considerably poorer κ values for both intra-observer and inter-observer tests on *all* categories than variables analysed by

assisted qualitative evaluation or by counting on-the-ball events. For fifteen variables evaluated only qualitatively, κ values for intra-observer agreement ranged from 0.94 to 0.59, while κ values for inter-observer agreement ranged from 0.82 to 0.19. For five variables analysed by assisted qualitative evaluation, κ values for intra-observer agreement ranged from 0.94 to 0.90, whereas κ values for inter-observer agreement ranged from 0.85 to 0.77. For two variables analysed by counting on-the-ball events, κ values for intra-observer agreement ranged from 0.94 to 0.92, whereas κ values for inter-observer agreement ranged from 0.90 to 0.84.

Table 6. Frequencies of team possession types (main offensive tactics) analyzed by the two observers (O1, O2) for inter- (O1-1, O2-1) and intra-observer (O1-1, O1-2) agreement test on *all* categories.

	Counter attack		Set play		Elaborate attack		Other		Total	
O1-1	O2-1	O1-2	O2-1	O1-2	O2-1	O1-2	O2-1	O1-2	O2-1	O1-2
Counter attack	39	56	0	0	3	2	8	0	50	58
Set play	0	0	33	36	2	1	2	0	37	37
Elaborate attack	20	4	14	0	64	98	3	0	101	102
Other	0	0	0	0	1	0	2	3	3	3
Total	59	60	47	36	70	101	15	3	191*	200

*Nine cases in which a particular type of category "Other" was registered by only one of the two observers were omitted from the test.

Table 7. Frequencies of defensive backup (opponent's defensive tactics) analyzed by the two observers (O1, O2) for inter-observer (O1-1, O2-1) agreement test on *all* categories.

	O2-1						Total
O1-1	Absent backup	More absent backup	Neutral backup	More present backup	Present backup	Not applicable	
Absent backup	48	32	20	6	20	1	127
More absent backup	1	9	7	6	1	0	24
Neutral backup	1	4	3	2	5	0	15
More present backup	0	0	0	1	1	0	2
Present backup	0	0	1	0	6	0	7
Not applicable	6	0	0	0	9	9	24
Total	56	45	31	15	42	10	199*

*One case in which category "Other" was registered by only one of the two observers were omitted from the test.

Table 8. Frequencies of skill level (secondary offensive tactics) analyzed by the two observers (O1, O2) for inter-observer (O1-1, O2-1) agreement test on *all* categories.

O1-1	O2-1			Total
	Neutral skill-level	More non-advanced skill	Non-advanced skill	
Neutral skill-level	1	1	0	2
More non-advanced skill	0	9	0	9
Non-advanced skill	12	33	131	176
Total	13	43	131	187*

*13 cases in which categories "Advanced skill" (7), "More advanced skill" (4), and "Not applicable" (2) were registered by only one of the two observers were omitted from the test.

Tables 3 through 6 show the frequency data in each category used by the two observers to compute inter- and intra-observer agreement (Table 6) for the main variable "team possession type" and the three variables with the lowest inter-observer reproducibility, namely "defensive backup" (poor), "skill level" (fair), and "defensive cover" (fair). For the variable "team possession type" which displayed moderate inter-observer reproducibility, 23 of 191 (12%) team possessions were characterized most differently by the two observers into *elaborate attack* and *counter attack* (Table 6). The same variable registered a very good intra-observer reproducibility with 6 of 200 (3%) team possessions characterized most differently by the same observer into *elaborate attack* and *counter attack* (Table 6). For the variable "defensive backup", 33 of 199 (17%) team possessions were characterized most differently into *absent backup* and *more absent backup* (Table 7).

Table 9. Frequencies of defensive cover (opponent's defensive tactics) analyzed by the two observers (O1, O2) for inter-observer (O1-1, O2-1) agreement test on *all* categories.

O1-1	O2-1					Total
	More absent cover	Neutral cover	More present cover	Present cover	Not applicable	
More absent cover	1	0	0	0	0	1
Neutral cover	0	2	1	3	0	6
More present cover	0	0	2	10	0	12
Present cover	0	3	7	146	1	157
Not applicable	0	0	0	15	9	24
Total	1	5	10	174	10	200

For the variable "skill level", 33 of 187 (18%) team possessions were characterized most differently into *non-advanced skill* and *more non-advanced skill* (Table 8). For the variable "defensive cover", 17 of 200 (9%) team possessions were characterized most differently into *present cover* and *more present cover* (Table 9).

4. Discussion

The main outcome of this study was that team match performance analysis using categorical data and opponent interaction has been developed as a reliable method to characterize team match performance in soccer for most variables used. Considering the challenges in describing different dimensions of soccer match performance with many dynamically interacting players by using videotapes recorded from a conventional TV coverage, the achieved reproducibility for most variables used was rather high. The results showing greater differences in inter-observer than intra-observer data were as expected.

It should be noted that the use of videotapes not specially filmed for the analysis of players' positional characteristics caused observational limitations which must be taken into account when interpreting the results of this study. Typical observational limitations experienced include difficulties in evaluation of players' positions and distances, and in determination of areas on the pitch. As TV production contained continuously varied camera angles and image sizes and the fact that there were no demarcation boundaries to show the area of the pitch in which an event took place, errors were likely to happen even with well prepared operational definitions and adopted practical measures. It is therefore not surprising that variables especially those involving evaluation of players' positional characteristics like "defensive backup" registered relatively poor reproducibility. Further studies are necessary using video material filmed from fixed optimal distances and angles in relation to the playing field. In this way match action can be captured in constant image sizes and with minimum parallax effect.

Further, this study has strengths worthy to be considered during interpretation of its results. It is based on team possessions randomly extracted from matches played in the Norwegian professional league. Therefore its results are representative of all matches at this level. Also, its sample size is estimated from the calculations involving each variable used and thus increasing the possibility to correctly assess their reliability. This means the presented results may be regarded as a genuine reflection of the way in which the subsequent data was analysed. Moreover, the study used Kappa statistics, an appropriate statistical method to test reliability of categorical data (Altman, 1991; Robinson & O'Donoghue, 2007). Both intra- and inter-observer agreement were tested, with adequate time of three weeks between test and retest to reduce problems of task familiarity. The comparison with earlier studies is difficult because the type of data and statistical methods employed vary between most studies. Only two previous studies used categorical data and Kappa statistics, and they report similar results with generally higher reproducibility than the current study. Bloomfield *et al*. (2005) used three variables of possession time, each with three categories, to measure intra- and inter-observer reliability and report κ values ranged between 0.81 to 0.98 and 0.72 to 0.92, respectively. Unfortunately, size of the sample and other relevant information about reliability procedures used were not reported. Seabra and Dantas (2006) used Space of Defensive Occupation (SDO zones) with seven categories of pass-reception situations to measure intra- and inter-observer agreement. Two experienced observers analysed 52 situations, starting after one week of analysis training and with three-week period between test and retest. κ values ranging from 0.83 to 0.93 for intra-reliability and from 0.73 to 0.90 for inter-reliability were reported.

The relatively low reproducibility reported especially for inter-observer reliability in the current study is most probably due to the complexity of its analysis system, with many analysis variables involving qualitative evaluation. In support, James, Taylor and Stanley (2007) argue that the disparity between two observers is to be expected especially when the analysis system involves considerable skill and experience. Hence, factors such as inadequate analysis training, inaccurate operational definitions and nature of analysis variables must be taken into account in the interpretation of its results. The four-week period used by the less experienced observer to train team match performance analysis may have been insufficient. The extent of difference in inter-observer data may reflect the considerable experience difference between the two observers. James *et al*. (2002) suggest that the variation between experienced and inexperienced observers found in their study was due to the relative lack of analysis training rather than inaccurate operational definitions. In their study, the experienced observer exhibited a 99% agreement for intra-observer reliability. For inter-observer reliability the two inexperienced observers registered relatively low (3%) and high (8%) levels of definition errors and observational errors, respectively. This suggests that the two-hour period they used to train inexperienced observers was inadequate, while the applied operational definitions were clear and easy to use. However, in our study, observation in 15 of 22 (68%) variables was done as a purely qualitative evaluation and achieving precise operational definitions for all variables used is difficult. O'Donoghue (2007) argues precise operational definitions do not guarantee good reliability and that good knowledge of the behaviours being analysed is more important than agreement of the wording of operational definitions. However, we cannot be certain since the effect of experience difference between the two observers was not measured in this study.

Some observations seem to be naturally more difficult to make without errors than others (James *et al*., 2007) and it is then logical to have different levels of reproducibility for different variables (Hughes *et al*., 2004). The current study revealed lower reproducibility for variables involving more qualitative evaluation than variables involving less or no qualitative evaluation for both intra-observer and inter-observer agreement. Indeed, all 10 variables with moderate or below κ values for the inter-observer agreement were analysed by qualitative evaluation alone. This included three variables with the lowest inter-observer reproducibility namely "defensive backup" (poor), "skill level" (fair), and "defensive cover" (fair). Moreover, somewhat lower reproducibility for variables involving identification of pitch areas (pitch-area variables) compared to variables of on-the-ball events (technical variables) found in our study corresponds to findings in earlier studies. In the present study, pitch-area variables "starting zone" and "starting corridor" had κ values of 0.92 and 0.90 for intra-observer agreement, while κ values for inter-observer agreement were 0.85 and 0.85, respectively. For technical variables "pass number" and "playing tempo", κ values for intra-observer agreement were 0.94 and 0.92, whereas κ values for inter-observer agreement were 0.84 and 0.90, respectively. Tucker *et al*. (2005) and Taylor *et al*. (2005) report errors of <5% in technical variables, whereas pitch-area variables recorded acceptable larger errors of <10% due to the identification difficulties mentioned above. Similarly, Hughes and Snook (2006) report that effort was needed to upgrade the reliability of pitch-area variables to <10% error limit. It seems that smaller differences between reproducibility of pitch-area and technical variables found in the current study than

those reported in the earlier studies is because of less identification difficulties experienced in our study (four and five pitch areas) compared to earlier studies (9, 18 and 36 pitch areas). These findings emphasize the need to consider differing levels of reliability for variables analysed by qualitative evaluation (Tucker *et al*., 2005).

The inspection of the tables of frequencies reveals that the two observers mostly failed to distinguish between two different offensive tactics, namely *elaborate attack* and *counter attack* in the main variable "team possession type". While for the three variables with the lowest inter-observer reproducibility, namely "defensive backup", "skill level" and "defensive cover", the two observers mostly failed to distinguish between two less different tactics. Thus, main sources of low reproducibility for the three variables with the lowest inter-observer reproducibility may be considered less serious compared to the main source of low inter-observer reproducibility for the main variable. This is because, for a Likert-type scale such as ordered categories in these cases, a disagreement between two less different tactics is considered as less serious than a disagreement between two more different tactics (Altman, 1991; Robinson & O'Donoghue, 2007). Indeed, grouping the data into few categories changed κ values for inter-observer reproducibility without losing practical meaning. For example, κ improved from 0.59 to 0.63 for "team possession type"; from 0.19 to 0.24 for "defensive backup"; from 0.25 to 0.26 for "skill level" and decreased from 0.35 to κ=0.27 for "defensive cover" (Table 5). This means that most of these variables, including the three variables with the lowest inter-observer reproducibility, can be meaningfully collapsed into fewer and more reliable categories, with acceptable reliability. Hence, these three variables can also be utilised in further analysis provided that care is taken when interpreting the data.

It is essential to ensure validity and adequate reliability of the collected data for match performance analysis to work effectively. It is important to realise that match performance is a product of the confrontation between two teams, involving multiple dimensions of performance (Harris & Reily, 1988; Grehaigne, 1991; Suzuki & Nishijima, 2004). Match performance is in most contemporary studies analysed by using unidimensional frequency data without considering opponent interaction. In our opinion, the analysis of match performance must include opposition relationship involving different dimensions of performance to be more valid. Team match performance analysis introduced by this study has been developed with this in mind - to analyse opposition relationship by using multidimensional qualitative evaluation through categorical variables representing offensive and defensive playing tactics, and their outcomes.

This new analysis system has the potential to evaluate and compare the effect of the interaction between different offensive tactics (e.g. counter attack vs. elaborate attack) and defensive tactics (e.g. imbalanced defence vs. balanced defence) on possession outcomes (e.g. shooting opportunity) in random series of team possessions. Also, causes as well as predictors of possession outcomes (e.g. goal scoring) can be determined from the interaction of playing tactics, both individually and as a group to evaluate also combining effect of different interrelated factors.

The effects of home versus away match location on match performance can also be assessed more effectively with this analysis system. Only few studies have documented the difference in technical and tactics-related behaviours as a function of match location in soccer (e.g. Sasaki *et al*., 1999; Tucker *et al*., 2005; Lago & Martin, 2007; Taylor *et al*., 2008). In addition, none of these studies has considered assessment of opponent interaction in their analyses.

Team match performance analysis can also be a useful tool in the evaluation of the measures of offensive effectiveness. Since goals provide few data points for the whole match (Harris & Reilly, 1988), the use of broader measures such as scoring opportunity (e.g. Olsen & Larsen, 1997), shot at goal (e.g. Pollard, 1986; Harris & Reilly, 1988; Hughes & Snook, 2006) and entry into final third (e.g. Bate, 1988) is common. The use of such broader measures may enable soccer practitioners to objectively see behind single match results, which are often influenced by chance. However, whether these measures are good enough and their relative ability to explain goal scoring has not been examined.

5. Conclusions

This study shows that a new method for team match performance analysis, which includes an assessment of opponent interaction and multidimensional categorical data, has been developed as a reliable method for most variables used. Thus, it represents a potentially valuable tool for more valid assessment of team match performance.

6.References

Altman, D. G. (1991). Some common problems in medical research. In Altman, D.G., .(Ed.), **Practical statistics for medical research,** pp. 403-409. London: Chapman & Hall.

Atkinson, G. & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports Medicine. **Sports Med,** 26 (4), 217-238.

Bate, R. (1988). Football chance: tactics and strategy. In T. Reilly, A. Lees, K. Davids & W. J. Murphy (Eds.), **Science and Football,** pp. 293-301. London: E & FN SPON.

Bloomfield, J. R., Polman, R. C. J., & O'Donoghue, P. G. (2005). Effects of score-line on team strategies in FA premier league soccer. **Journal of Sports Sciences,** 22, 505.

Cooper, S. M., Hughes, M., O'Donoghue, P., & Nevill, A. M. (2007). A simple statistical method for assessing the reliability of data entered into sport performance analysis systems. **International Journal of Performance Analysis in Sport,** 7(1), 87-109.

Elias, N. & Dunning, E. (1966). Dynamics of group sports with special references to football. **British Journal of Sociology,** 17, 388-402.

Grehaigne, J. F. (1991). A new method of goal analysis. **Science and Football,** 5, 10-16.

Grehaigne, J. F. & Godbout, P. (1995). Tactical knowledge in team sports from a constructivist and cognitivist perspective. **Quest,** 47 (4), 490-505.

Grehaigne, J. F., Bouthier, D. & David, B. (1997). Dynamic-system analysis of opponent .relationships in collective actions in soccer. **Journal of Sports Sciences,** 15, 137-149.

Grehaigne, J. F., Bouthier, D. & Godbout, P. (1999). The foundations of tactics and strategy in team sports. **Journal of Teaching in Physical Education,** 18 (2), 159-174.

Grehaigne, J. F., Mahut, B., Fernandez, A. (2001). Qualitative observation tools to analyse soccer. **International Journal of Performance Analysis in Sport,** 1(1), 52-61.

Harris, S. and Reilly, T. (1988). Space, team work and attacking success in soccer. In T. Reilly, A. Lees, K. Davids & W. J. Murphy (Eds.), **Science and Football,** pp. 322-328. London: E & FN SPON.

Hughes, M. D. & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. **Journal of Sports Sciences,** 20, 739-754.

Hughes, M. & Churchill, S. (2004). Attacking profiles of successful and unsuccessful teams in Copa America 2001. **Journal of Sports Sciences,** 22, 505.

Hughes, M. & Reed, D. (2005). Creating a performance profile using perturbations in soccer. In D. Milanovic & F. Prot (Eds.), **Proceedings of 4th International Scientific Conference on Kinesiology,** pp. 34-53. Zagreb: University of Zagreb.

Hughes, M. & Snook, N. (2006). Effectiveness of attacking play in the 2004 European Championships: In H. Dancs, M. Hughes & P. G. O'Donoghue (Eds.), **Proceedings of World Congress of Performance Analysis of Sport 7,** pp. 46-62. Cardiff: CPA UWIC Press.

Hughes, M. Dawkins, N., Reed, D. & Mills, J. (1998). The perturbation effect and goal opportunities in soccer. **Journal of Sports Sciences,** 16, 20.

Hughes, M., Cooper, S. M. & Nevill, A. (2004). Analysis of notation data – reliability. In M. Hughes & I. M. Franks (Eds.), **Notational Analysis of Sport,** pp. 189-204. London: Routledge.

James, N., Mellalieu, S. D. & Hollely, C. (2002). Analysis of strategies in soccer as a function of European and domestic competition. **International Journal of Performance Analysis in Sport,** 2(1), 85-103.

James, N., Taylor, J., & Stanley, S. (2007). Reliability procedures for categorical data in performance analysis. **International Journal of Performance Analysis in Sport,** 7(1), 1-11.

Jones, P. D., James, N. & Mellalieu, S. D. (2004). Possession as a performance indicator in soccer. **International Journal of Performance Analysis in Sport,** 4(1), 98-102.

Konstadinidou, X. & Tsigilis, N. (2005). Offensive playing profiles of football teams from the 1999 Women's World Cup Finals. **International Journal of Performance Analysis in Sport,** 5(1), 61-71.

Lago, C. & Martin, R. (2007). Determinants of possession of the ball in soccer. **Journal of Sports Sciences,** 25, 969-974.

Luhtanen, P., Belinskij, A., Hayrinen, M., & Vanttinen, T. (2001). A comparative tournament analysis between the EURO 1996 and 2000 in soccer. **International Journal of Performance Analysis in Sport,** 1(1), 74-82.

McGarry, T. & Franks, I. M. (2003). The science of match analysis. In T. Reilly & A. M. Williams (Eds.), **Science and Soccer** (2nd edition), pp. 265-275. London: Routledge.

O'Donoghue, P. (2007). Reliability issues in performance analysis. **International Journal of Performance Analysis in Sport,** 7(1), 35-48.

Olsen, E. & Larsen, O. Use of match analysis by coaches. In T. Reilly, J. Bangsbo & M. Hughes (Eds.), **Science and Football III,** pp. 209-220. London: E & FN SPON.

Pollard, R. (1986). Soccer performance analysis and its application to shots at goal.

Research Bi-annual for Movement, 4, 19-27.
Pollard, R., & Reep, C. (1997). Measuring the effectiveness of playing strategies at soccer. **The Statistician,** 46, 541-550.
Robinson, G. & O'Donoghue, P. (2007). A weighted Kappa statistic for reliability testing in performance analysis of sport. **International Journal of Performance Analysis in Sport,** 7(1), 12-19.
Sasaki, Y., Nevill, A. & Reilly, T. (1999). Home advantage – A case study of Ipswich Town football club during the 1996-97 season. **Journal of Sports Sciences,** 17, 831.
Scoulding, A., James, N., & Taylor, J. (2004). Passing in the soccer World cup 2002. **International Journal of Performance Analysis in Sport,** 4(2), 36-41.
Seabra, F. & Dantas, L. (2006). Space definition for match analysis in soccer: In H. Dancs, M. Hughes & P. G. O'Donoghue (Eds.), **Proceedings of World Congress of Performance Analysis of Sport 7,** pp. 30-45. Cardiff: CPA UWIC Press.
Suzuki, K. & Nishijima, T. (2004). Validity of a soccer defending skill scale (SDSS) using game performances. **International Journal of Sport and Health Science,** 2, 34-49.
Taylor, J. B., Mellalieu, S. D. & James, N. (2004). Behavioural comparisons of positional demands in professional soccer. **International Journal of Performance Analysis in Sport,** 4(1), 81-97.
Taylor, J. B., Mellalieu, S. D. & James, N. (2005). A comparison of individual and unit tactical behaviour and team strategy in professional soccer. **International Journal of Performance Analysis in Sport,** 5(2), 87-101.
Taylor, J. B., Mellalieu, S. D., James, N. & Shearer, D. A. (2008). The influence of match location, quality of opposition, and match status on technical performance in professional association football. *Journal of Sports Sciences,* 26, 885-895.
Tucker, W., Mellalieu, S. D., James, N. & Taylor, J. B. (2005). Game location effects in professional soccer: A case study. **International Journal of Performance Analysis in Sport,** 5(2), 23-35.
Worthington, E. (1980). **Teaching soccer skills.** London: Henry Kimpton (Publishers) Ltd.