Journal of Sports Sciences
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rjsp20

Analysis of passing sequences, shots and goals in soccer

Mike Hughes [a] & Ian Franks [b]

[a] Centre for Performance Analysis, University of Wales Institute Cardiff, Cyncoed, Cardiff, UK

[b] School of Human Kinetics, University of British Columbia, Vancouver, Canada

[c] Centre for Performance Analysis, University of Wales Institute Cardiff, Cyncoed, Cardiff, CF2 6XD, UK E-mail:

Version of record first published: 18 Feb 2007.

To cite this article: Mike Hughes & Ian Franks (2005): Analysis of passing sequences, shots and goals in soccer, Journal of Sports Sciences, 23:5, 509-514

To link to this article: http://dx.doi.org/10.1080/02640410410001716779

Journal of Sports Sciences, May 2005; 23(5): 509 – 514

Taylor & Francis
Taylor & Francis Group

Analysis of passing sequences, shots and goals in soccer

MIKE HUGHES[1] & IAN FRANKS[2]

[1]*Centre for Performance Analysis, University of Wales Institute Cardiff, Cyncoed, Cardiff, UK, &* [2]*School of Human Kinetics, University of British Columbia, Vancouver, Canada*

(Accepted 5 February 2004)

Abstract
Early research into how goals were scored in association football (Reep and Benjamin, 1968) may have shaped the tactics of British football. Most coaches have been affected, to a greater or lesser extent, by the tactics referred to as the ''long-ball game'' or ''direct play'', which was a tactic employed as a consequence of this research. Data from these studies, published in the late 1960s, have been reconfirmed by analyses of different FIFA World Cup tournaments by several different research groups. In the present study, the number of passes that led to goals scored in two FIFA World Cup finals were analysed. The results conform to that of previous research, but when these data were normalized with respect to the frequency of the respective lengths of passing sequences, there were more goals scored from longer passing sequences than from shorter passing sequences. Teams produced significantly more shots per possession for these longer passing sequences, but the strike ratio of goals from shots is better for ''direct play'' than for ''possession play''. Finally, an analysis of the shooting data for successful and unsuccessful teams for different lengths of passing sequences in the 1990 FIFA World Cup finals indicated that, for successful teams, longer passing sequences produced more goals per possession than shorter passing sequences. For unsuccessful teams, neither tactic had a clear advantage. It was further concluded that the original work of Reep and Benjamin (1968), although a key landmark in football analysis, led only to a partial understanding of the phenomenon that was investigated.

Keywords: *Analysis, association football, goals, passing sequence, shots, soccer*

Introduction

The tactics and pace of play by teams in British association football sets it apart from soccer played by teams in Europe and other continents. This was demonstrated by Yamanaka *et al.* (1993) at the most elite level, the FIFA 1990 World Cup, but it would also seem to apply at most levels of club performance. One of the reasons for this polarization of tactics could be traced back to events following match analysis performed by Reep, who collected data from 3213 matches between 1953 and 1968 (Reep & Benjamin, 1968). These data on goal-scoring and the length of passing sequences were analysed statistically and presented as a negative binomial distribution. This idea was then extended to other sports (Reep *et al.*, 1971), although no practical applications of these models were made to the respective sports.

Two main findings arose from the analyses of Reep and colleagues: (1) approximately 80% of goals resulted from a sequence of three passes or less and (2) a goal is scored every 10 shots. These findings have been confirmed by several research groups analysing the games of different FIFA World Cup finals (Franks *et al.*, 1983, 1990; Hughes *et al.*, 1988; Partridge & Franks, 1989a,b; Grehaigne, 1999). Several coaches used these results to evolve a simple tactical approach to soccer, which was to maximize the ''chance'' elements of the game in favour of their teams. Briefly, the tactical implication was for the team in possession to move the ball into a shooting position as directly as possible with the least number of passes (see Franks, 1996, for a detailed description and critical discussion of the tactic termed ''direct play''). Bate (1988) extended some of these ideas by exploring aspects of chance in football and its relation to tactics and strategy in light of the results presented by Reep and Benjamin (1968) and data from unpublished research collected by Hughes in 1987. Bate (1988) rejected the concept of ''possession football'' and advocated a more direct strategy with fewer passes per team possession. As might be expected, these tactics were not without controversy

Correspondence: Mike Hughes, Centre for Performance Analysis, University of Wales Institute Cardiff, Cyncoed, Cardiff CF2 6XD, UK.
E-mail: mhughes@uwic.ac.uk

ISSN 0264-0414 print/ISSN 1466-447X online © 2005 Taylor & Francis Group Ltd
DOI: 10.1080/02640410410001716779

in both coaching and research communities (see Franks, 1989, for an interesting debate on the issue).

Several teams have achieved a measure of success using these strategies, particularly in progressing from the lower divisions of English football (e.g. Wimbledon and Watford). Indeed, Eire and Norway are regarded as over-achievers in the international arena, using low passing sequences per possession with very limited playing resources. However, very few teams have succeeded at the highest level by winning the World Cup or other European championships using the tactic ''direct play''. Nevertheless, the effect of these tactics did permeate the manner in which most clubs played in Britain, and still do.

The fact that successful (league champions, World champions, European champions) teams did not use ''direct play'' would indicate that there are further dimensions of these types of data to be explored. Hughes *et al.* (1988) examined patterns of play for successful (semi-finalists) and unsuccessful (eliminated at end of first round) teams in the 1986 World Cup finals. These authors found that successful teams played significantly more touches of the ball per possession than unsuccessful teams. The interpretation of the empirical model posed by Reep and Benjamin (1968), by researchers and coaches (e.g. Hughes, 1987; Bate, 1988; Franks, 1989), did not appear to fit the performance characteristics at all levels of soccer.

It is proposed that the accepted fact that ''80% of goals resulted from a sequence of three passes or less'' may be misinterpreted. In branches of mathematics, when treating unequal frequencies of occurrences, the outcomes are ''normalized'' by dividing the number of outcomes by the frequency of their occurrences. These ideas have not been applied generally to performance analysis data. With some relatively large databases (the 1990 and 1994 World Cups) available for analysis, it was now possible to re-analyse goals, shots and lengths of possession. The purpose of this paper is to normalize these data to enable an examination of different interpretations, specifically comparing successful and unsuccessful teams.

Methods

Databases of the 1990 (Italy) and 1994 (USA) FIFA World Cup finals for soccer were analysed. Both of these competitions had been notated by two research groups, one in the Centre for Notational Analysis, Liverpool John Moores University, UK and the other in the Human Kinetics Laboratory, University of British Columbia, Vancouver, Canada; both notation systems were independently tested for validity and reliability (Partridge & Franks, 1989a; Yamanaka *et al.*, 1993). Each system offered different aspects of the data from a soccer match – hence the use of both databases. The databases were further cross-checked for inter-system accuracy, and simple percentage agreements greater than 99% were obtained for shots, goals and passes per possession.

A passing sequence for one team possession was defined in terms of sequence length. A sequence length of zero was an intended pass that was contacted by the opposition. A sequence length of 1 was an intended pass that was contacted by fellow team-mates but then the second pass was contacted by the opposition. A two-pass sequence was terminated when the third intended pass did not reach its target, and so on.

The 1990 World Cup involved 24 teams and a total of 52 matches, whereas the 1994 format was expanded to 32 teams and a corresponding total of 64 matches.

Results and discussion

Goal-scoring

The data collected on goals scored for each possession ''passing sequence'' (see Figure 1) were in agreement with the results of Reep and Benjamin (1968). In 1990, 84% of goals came from team possessions of four passes or less, whereas in 1994 the figure was 80%. Nevertheless, there is some variance in the data as shown by the function of the graph in Figure 1, with longer passing sequences having lower frequencies of goals. Furthermore, there were differences in the numbers of teams (16 and 24 respectively) and, therefore, the numbers of matches (52 and 64 respectively) at the two tournaments.

The nature of the distribution of the lengths of passing sequences, shown in Figure 2, suggests that the incontrovertible fact that approximately 80% of goals are scored from a total of four passes or less (as observed in Figure 1), may be interpreted in other ways. There are higher frequencies of possessions of

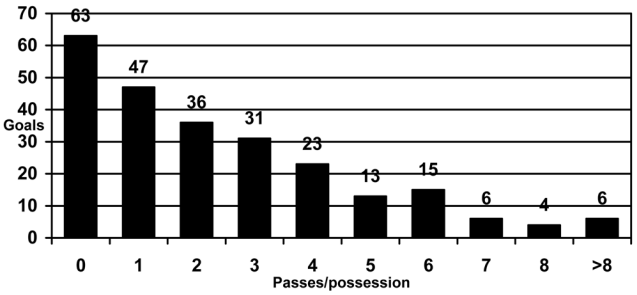

Figure 1. Patterns of goal-scoring with respect to the different lengths of possessions in the 1990 and 1994 World Cups for soccer.

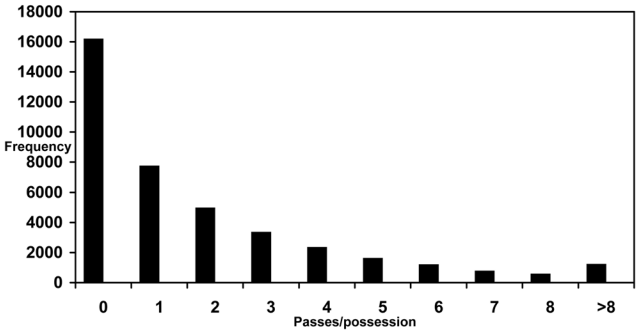

Figure 2. Frequency of each sequence length in the 1990 and 1994 World Cups for soccer.

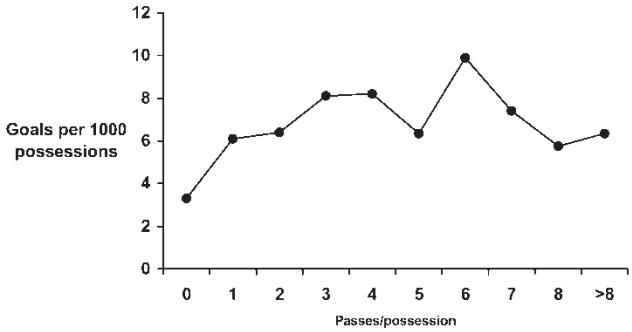

Figure 3. Analysis of the mean number of goals scored per 1000 possessions for the 1990 and 1994 World Cups.

short passing sequences than long sequences of passing. Is the scoring of goals, or indeed shooting, dependent upon the length of passes per possession?

To analyse the relative importance of the conversion rates from the different lengths of passing sequences per possession into goals, it is necessary to assess the relative contribution of each possession length from equal frequencies of occurrences. This is achieved by normalizing the data for the purpose of comparisons. In the present study, the data were normalized by dividing the number of goals scored in each team possession by the frequency of that sequence length. To avoid very small ratios, the data were presented as ''goals per 1000 possessions'' for each respective possession length (see Figure 3). A different profile of the relative importance of the different sequence lengths is obtained. The curve shows that the longer sequence lengths have a higher conversion ratio of goals per 1000 possessions. This would indicate that teams that have the skill to sustain long passing sequences have a better chance of scoring. This simple analysis of one variable shows that there may be other interpretations of Reep and Benjamin's (1968) data.

The maximum number of passes per possession in the 1990 World Cup is approximately seven, while the 1994 World Cup shows a peak in the data at about six passes per possession (see Figure 4). Although this difference in the data may be due to sampling errors, one could speculate that these differences are the result of individual variance (i.e. the teams involved) and/or environmental differences in the two tournaments. Future research could try to profile individual countries and, in particular, contrast the British teams with other top-class teams such as Brazil, Germany and Italy.

Shooting and lengths of passing sequences per possession

Because there are approximately 10 shots for every goal, similar analyses for shots could reveal more information about the process of goal scoring. The

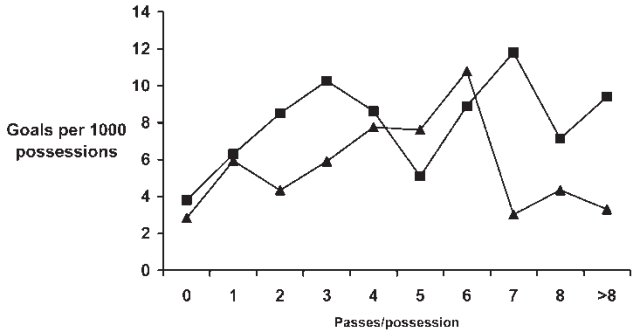

Figure 4. Analysis of the number of goals scored per 1000 possessions for the 1990 (■) and 1994 (▲) World Cups.

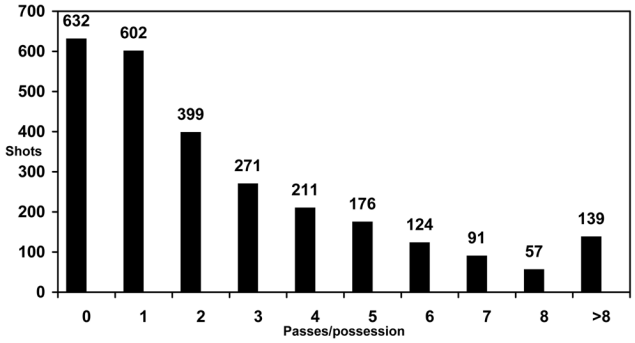

Figure 5. Frequency of shots with respect to length of possession in the two World Cup tournaments.

frequency of shots with respect to the length of passing sequences gave a similar profile to that obtained for goals, with 80% of the shots in 1990 occurring for passing sequences of four or less; the corresponding figure for 1994 was 77% (Figure 5).

Normalizing these data with respect to the frequency of the passing sequences gives the data distribution shown in Figure 6, which is similar in shape to the distribution for goals (Figures 3 and 4). However, the 1990 data peak at a possession length of seven, while the 1994 data peak at a value of four, the mean peaking at five passes per possession. The

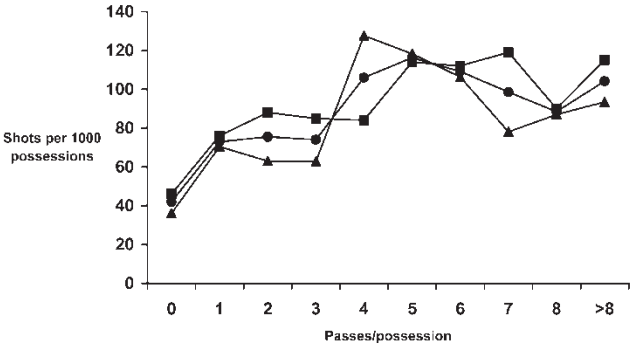

Figure 6. Frequency of shots per 1000 possessions of different length for the 1990 (■) and 1994 (▲)World Cups. ● = mean.

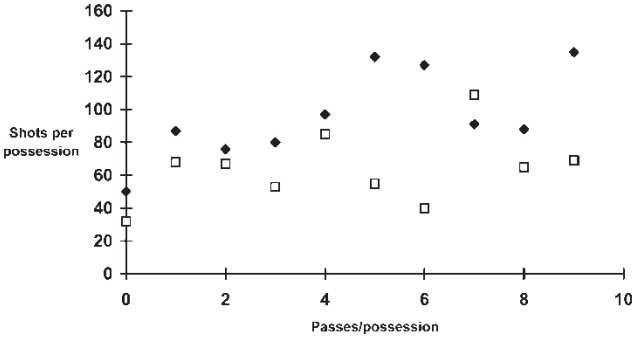

Figure 7. The mean numbers of shots per 1000 possessions for successful (◆) and unsuccessful (□) teams in the 1990 World Cup.

Table I. The conversion ratios of shots to goals for the 1990 and 1994 World Cups with respect to the length of possession

	Number of shots per goal per 1000 possessions		
Possessions	*1990*	*1994*	*Mean*
0	9.1	8.9	9.0
1	12.6	9.2	11.9
2	8.4	14.5	11.5
3	7.1	10.4	8.8
4	6.5	7.8	7.1
Mean 0–4 passes/possession	8.7	10.2	9.5
5	19.8	10.8	15.3
6	9.3	7.5	8.4
7	9.0	28.0	18.5
8	14.0	14.5	14.3
>8	12.0	25.7	18.9
Mean 5–8 passes/possession	12.8	17.3	15.1

longer passing sequences have higher frequencies of shots per 1000 possessions.

The data presented in Table I show the number of shots required to produce a goal at each of the different passing sequence lengths (normalized to 1000 possessions). The overall mean of the short passing sequences was very close to the shot to goal ratio found by Reep and Benjamin (1968). Using *t*-tests, the means of the conversion rates of possession lengths 0–4 were compared with the means of possession lengths 5–8; those of the longer passing sequences were significantly higher ($t_6 = -2.59$; $P < 0.05$; $\omega^2 = 0.27$; % increase = 56% between the two means of the mean column). The difference becomes significant for the means as the averaging process lessens the variance of the data. These trends and differences in the data and their means are thought to be due to the effectiveness of the quicker attack disrupting the defence. As a longer possession attack builds up, the defence has more opportunity to minimize surprise and dislocation of the defence by the attack.

If teams have the skill to sustain possession, they have a greater chance of creating a shooting opportunity, but the conversion ratios of shots to goals are lower at longer lengths of possession. Is it better, then, to use short passing sequences (direct play), given the better conversion ratio of shots to goals or longer possessions given that there is more likelihood of producing a shot?

Further analyses of the 1990 World Cup data on a team basis showed some interesting trends. Eight quarter-finalists were categorized as successful teams (England, Italy, Germany, Argentina, Eire, Cameroon, Czechoslovakia and Russia) and eight first round losers were deemed unsuccessful (Austria, Scotland, Egypt, USA, Costa Rica, Korea, Sweden and the United Arab Emirates). The shooting data of these teams were analysed for each of the different lengths of possession.

Individual trends were difficult to identify, but when the data for each team in the respective groups are added, and means for each possession string calculated, then there are some clearer messages (Figure 7). The data looked as though they could be linear or perhaps polynomial, so both sets were simulated using a computer graphics package (Figure 8). A regression analysis of the two sets of data showed that the data for the successful teams are a far better approximation to a straight line ($r = 0.69$) than the data for the unsuccessful teams ($r = 0.38$).

The data show similar trends to the analyses of the different rounds, with the successful and unsuccessful teams having similar shooting figures at lower passing possessions, but these data diverge at the longer possessions. A synopsis of these data is given in Table II, differentiating between the numbers of shots for the successful and the unsuccessful teams at different lengths in possession. A *t*-test showed that the two groups were significantly different in producing shots at both short and long possessions ($P < 0.05$). The difference between the number of

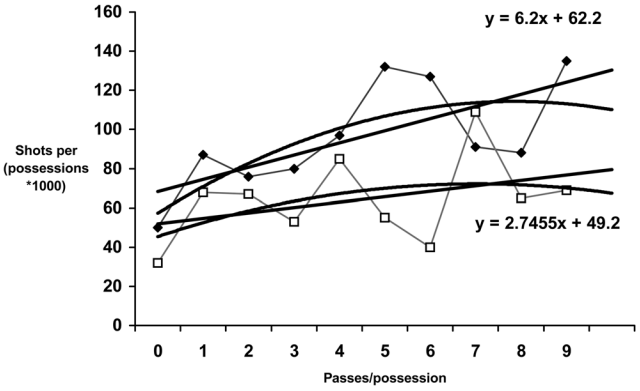

Figure 8. The mean numbers of shots per 1000 possessions for successful (◆) and unsuccessful (■) teams, with different graphical interpretations.

shots per 1000 possessions at short and long possessions for the successful teams was significantly different ($P<0.01$), but there was no significance in the test for unsuccessful teams. If teams have the skill to sustain possession, they stand a far greater chance of creating a shooting opportunity, but again the conversion ratios of shots to goals are lower at these longer lengths of possession. Is it better, then, to use short direct possessions, given the better conversion ratio of shots to goals or longer possessions given that there is more likelihood of producing a shot? Attempting to combine the data on shots per possession with the ratios of conversion could give an estimate of the relative values of using short or long passing attacks on goals. This is attempted in Table II and the data show that, although the shots to goal ratio is lower for these longer possessions, in balance they are a more efficient way of scoring goals. The problem for coaches is the choice of balancing the possession football with incisive, higher risk play.

It is accepted that the more successful teams will have better conversion ratios than the unsuccessful teams, but even without these conversion ratio data it can be seen that the longer possessions can and should yield a better reward. The difficulty facing coaches is to determine the skill of their team and then the appropriate tactics, but it would seem that the principle of direct play is only applicable where the skill of the team is insufficient to sustain possession of the ball.

The interesting fact about these results for the performance analyst is that although the data presented by Reep and Benjamin (1968; Reep *et al.*, 1971) have been replicated, different interpretations can be obtained from the same data using different analyses. The analyses used in the present paper appear to be appropriate for these types of data. They do not refute the work of Reep and Benjamin; indeed, this work confirms their analyses at these different levels of soccer, but it is also apparent that further work is required to evaluate the processes used in performance analysis of team games.

Table II. Summary of the goal-scoring data for successful and unsuccessful teams per 1000 possessions in the 1990 World Cup

	Successful			*Unsuccessful*		
Possession length	*Shots*	*Shots/ goal*	*Projected goals*	*Shots*	*Shots/ goal*	*Projected goals*
0	70	9.4	7.4	50	9.4	5.3
> 8	125	14.8	8.5	75	14.8	5.1

Conclusions

The data presented here replicate those of Reep and Benjamin (1968), but when these same data were normalized with respect to the frequency of the respective lengths of possession, we found the following:

- There were significantly more shots per possession at longer passing sequences than there were at shorter passing sequences for successful teams.
- The conversion ratio of shots to goals is better for direct play than for possession play.
- A comparison of successful and unsuccessful teams in the 1990 World Cup showed differences between the two in converting possession into shots on goal, with the successful teams having the better ratios. These differences were not significant but indicated strong trends in the data.

It was further concluded that the manner in which the data were presented by Reep and Benjamin (1968) has lead to only a partial understanding of the phenomenon under investigation. This type of simple presentation is common in many publications on performance analysis. We strongly recommend more consideration should be given to normalizing frequency data with respect to the relevant total data sets (see Hughes and Bartlett, 2002). Future studies should extend this work by analysing greater quantities of data, and also establishing that the amount of data acquired is sufficient to definitely establish a performance profile, as some recent work has demonstrated that shooting and goal scoring stabilize with fewer matches than might be expected (Hughes *et al.*, 2001, 2003).

References

Bate, R. (1988). Football chance: tactics and strategy. In *Science and Football* (edited by T. Reilly, A. Lees, K. Davids and W. Murphy), pp. 293–301. London: E & FN Spon.

Franks, I.M. (1989). Critique but critique accurately and with the facts: a reply to Allen Wade. *Soccer Journal*, May–June, pp. 39–41.

Franks, I.M. (1996). Analysis of Association Football. In *Coaching Soccer* (edited by T. Schum), pp. 29–37. Indianapolis, IN: Master Press.

Franks, I.M., Goodman, D., & Miller, G. (1983). Human factors in sport systems. In *Proceedings of the Human Factors Society 27th Annual Meeting* (Norfolk, Virginia, October 10–14), pp. 383–386. Santa Monica, CA: Human Factors Society.

Franks, I.M., Partridge, D., & Nagelkerke, P. (1990). *World Cup 90: A Computer Assisted Technical Analysis of Team Performance*. Technical Report for the Canadian Soccer Association.

Grehaigne, J.-F. (1999). Systemic approach and soccer. In *Notation of Sport III* (edited by M. Hughes), pp. 1–8. Cardiff: Centre for Performance Analysis, UWIC.

Hughes, C. (1987). *The Football Association Coaching Book of Soccer Tactics and Skills*. London: Queen Anne Press.

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

Hughes, M.D., Robertson, K., & Nicholson, A. (1988). An analysis of the 1984 World Cup of Association Football. In *Science and Football* (edited by T. Reilly, A. Lees, K. Davids and W. Murphy), pp. 363–367. London: E & FN Spon.

Hughes, M., Evans, S., & Wells, J. (2001). Establishing normative profiles in performance analysis. *eIJPAS – International Journal of Performance Analysis in Sport* (*Electronic*), *1*, 4–27.

Hughes, M., Cooper, S.-M., Nevill, A., & Brown, S. (2003). An example of reliability testing and establishing performance profiles for non-parametric data from performance analysis. *International Journal of Computers in Sports Science*, *2*, 34–56.

Partridge, D., & Franks, I.M. (1989a). A detailed analysis of crossing opportunities from the 1986 World Cup (Part I). *Soccer Journal*, May–June, pp. 47–50.

Partridge, D., & Franks, I.M. (1989b). A detailed analysis of crossing opportunities from the 1986 World Cup (Part II). *Soccer Journal*, June–July, pp. 45–48.

Reep, C., & Benjamin, B. (1968). Skill and chance in association football. *Journal of the Royal Statistical Society*, *A*, *131*, 581–585.

Reep, C., Pollard, R., & Benjamin, B. (1971). Skill and chance in ball games. *Journal of the Royal Statistical Society*, *A*, *134*, 623–629.

Yamanaka, K., Hughes, M., & Lott, M. (1993). An analysis of playing patterns in the 1990 World Cup for association football. In *Science and Football* (edited by T. Reilly, A. Lees, K. Davids and W. Murphy), pp. 206–214. London: E & FN Spon.