

CHAPTER SEVEN

WRITTEN PERCEPTIONS OF REFEREEING PERFORMANCE

Introduction

Findings in the previous chapter provided substantial direction and focus for improving refereeing standards. However, the results were constrained somewhat due to the 'forced-choice' response options in Section 2 of the survey. This chapter addresses this issue by analysing related written comments supplied by respondents to Section 3 of the survey. Subsequent findings allow an overall evaluative impression of refereeing standards to be formed, and permits a more detailed examination of specific competencies.

To facilitate discussion, this chapter is divided into four sections. The first section describes the written comments, with particular reference to the compilation and classification of responses derived from the data analysis. The second section gives an insight into the perceptions of soccer stakeholders concerning the current refereeing standard in the Ericsson Cup. Following from this discussion, the third section provides a more fine-grained analysis of the data, and scrutinises comments related to specific competencies. The final section draws together substantive issues from the previous two sections, and links the qualitative evidence with the relevant quantitative results presented in Chapter 6.

OVERVIEW OF DATA

This section explores the nature of the written comments provided by respondents. Specifically, two aspects of the data are examined. These are: response-rates; and, a review of the data categorisation process, including the type and form of responses.

Response-Rates

Of the 173 respondents who completed the survey instrument, 118 provided responses in Section 3 of the instrument. This section comprised one open-ended question, namely, "Please provide your opinions on the current standard of refereeing in the Ericsson Cup." Subsequent response-rates, for each sub-group of stakeholders, are provided in Table 7.1.

Table 7.1 Percentage and frequency response rates by sub-group

	Response to Question	Total sub-sample	Percentage response
Referees	11	13	85
Assistant referees	24	26	92
Referee inspectors	23	24	96
Players	42	92	46
Coaches	18	18	100
Totals	118	173	68

The figures contained in Table 7.1 present an overall response-rate of 68%. This is acceptable given the survey question was optional. Specifically, all groups (with the exception of players) provided response-rates greater than 50% (46% for players, 85-100% for the remaining groups). A lower response-rate from players was anticipated given the indifference of many players toward administrative-type tasks (see Chapter 4). Nonetheless, the overall response-rate to the question ensured all sub-groups were represented adequately, and confirmed the value of the question to data analysis.

Categorisation of Data

Written comments were supplied in an unstructured format. To derive meaning from responses, the comments were analysed for common themes and issues using the QSR NUD•IST software package. As described in Chapter 4, this software facilitates the analysis of non-numerical data. Moreover, within the context of the present study, NUD•IST provides a mechanism to establish links between the written comments and other soccer-related information collected from the questionnaire, e.g., relative playing experience, state of residence, age, etc. These links facilitate a more ‘fine-grained’ analysis of the data.

Three aspects of analysis require clarification in this discussion. These are: the reliability of data categorisation process; the NUD•IST tree structure that resulted from the data classification; and, the varying quantity of comment provided by respondents.

Coding Reliability

The coding process involved all the written responses being entered into the NUD•IST software package as described in the NUD•IST protocol (QSR, 1996). However, the coding process raises concerns associated with inter- and intracoder reliability.

Specifically, intercoder reliability refers to the ability of two or more researchers to achieve the same or similar coding results from a set of responses. One measure of intercoder

reliability (Miles & Huberman, 1994, p. 64) is to express the number of agreements as a percentage of the number of coded items, i.e.,

$$\text{intercoder reliability} = \frac{\text{number of agreements}}{\text{total number of agreements} + \text{disagreements}}$$

When content analysis techniques are being used, it is reported that agreement measures as high as 90% can be achieved (Miles & Huberman, 1994). However, variability in agreement is not uncommon, and is dependent on the type of material being coded. For example, Kerlinger (1986) reported agreement coefficients of 70-80% being achieved in judging the creativity of student essays. Under conditions set by the present study, where the unit of analysis can be as small as one word, yet as large as 400 words, agreement coefficients greater than 80% could be expected. As a consequence, a conservative intercoder reliability coefficient of 85% was targeted for the present study.

To achieve this end, the researcher trained one assistant in coding procedures and category definitions. The assistant was asked to code a random sample of 40 responses. The codings were checked against the investigator's codings, and similarities and disagreements were discussed. One difference arose at this stage, i.e., a lack of clarity in category definitions. This resulted in definitions being amended by consensus between the investigator and the assistant. At the conclusion of this process, an inter-rater reliability coefficient of 91% was achieved. This represented an acceptable level of agreement.

With respect to intracoder reliability, it was important for the coder to maintain consistency in coding across time. Two threats emerged here. Firstly, subtle changes can occur in meaning, in the mind of the coder, of both coding definitions and subject responses. This phenomena, referred to as "retrospective hindsight" (Miles & Huberman, 1984, p. 57), tends to occur when the coder becomes familiar with the data. To minimise this threat, (while, at the same time, being cognisant of the need to code data accurately) the coding process was not prolonged unnecessarily. Additionally, the rigorous definitions applied to each classification of responses, refined in conjunction with the assistant, facilitated this endeavour.

The second threat arose from errors on the part of the investigator. To monitor and minimise this threat, an intracoder reliability co-efficient was calculated using the same formula as for determining the intercoder reliability co-efficient. Intracoder checks were done on three occasions by randomly selecting 10% of the investigators coded responses, recoding them and performing the necessary calculations. Calculations showed that differences occurred in fewer than 2% of cases. This small variation showed that errors emanating from investigator error were relatively few.

Classification Structure

Coding derived a substantive tree structure (see Figure 7.1)¹. However, before discussing the specific details of the tree, an explanation of how NUD•IST nodes are expressed, and the relationships among nodes, is pertinent (illustrative examples are given with reference to Figure 7.1).

Primarily, the tree structure is analogous to a family tree. Any node connected to a given node from below is a ‘child’ (e.g., *Communication* is a child of *Comments*), while a node connected from a node above is its ‘parent’ (e.g., *Base Data* is a parent of *Age*). A node can have any number of children, but only one parent. All children from the same parent are said to be each other’s ‘sibling’ (e.g., *Consistency* and *Communication* are siblings).

Importantly, there is always one node without siblings or parents. This is called the ‘root’ node, and all subsequent nodes emanate from this node. Lastly, any node in a tree, taken together with all its children, is said to be a ‘sub-tree.’ The node at the top of that tree is called a ‘root’ also. All nodes below it are ‘descendants.’ Similarly, the parent of a node, its parents and so forth are ‘ancestors’ of a given node (QSR, 1996).

The tree structure in Figure 7.1 comprised a root node with two children (*Base Data* and *Comments*). Primarily, the *Base Data* node contained all data collected from Section 1 of the survey, and the *Comments* node encompassed data from Section 3 of the survey. Subsequently, these nodes formed two sub-trees containing numerous children.

Specifically, children of the *Comments* node reflect diversity in the content respondents commented upon. The diversity of topics reinforce the concept of refereeing being a multi-faceted occupation, with many issues impacting on the quality of refereeing performance. A more detailed analysis of comments is provided later in this chapter.

With respect to children of the *Base Data* node, pertinent information within these nodes has been discussed previously in Chapter 5. However, the relevance of this information should not be overlooked, as the context of this analysis requires that information within these nodes be cross-referenced with *Comments* children.

Quantity of Comment

The analysis of written comments highlighted a large difference in the amount of comment provided by each respondent. Responses range from singular words and phrases, to in-

¹ Working definitions for each node are provided in Appendix 26.

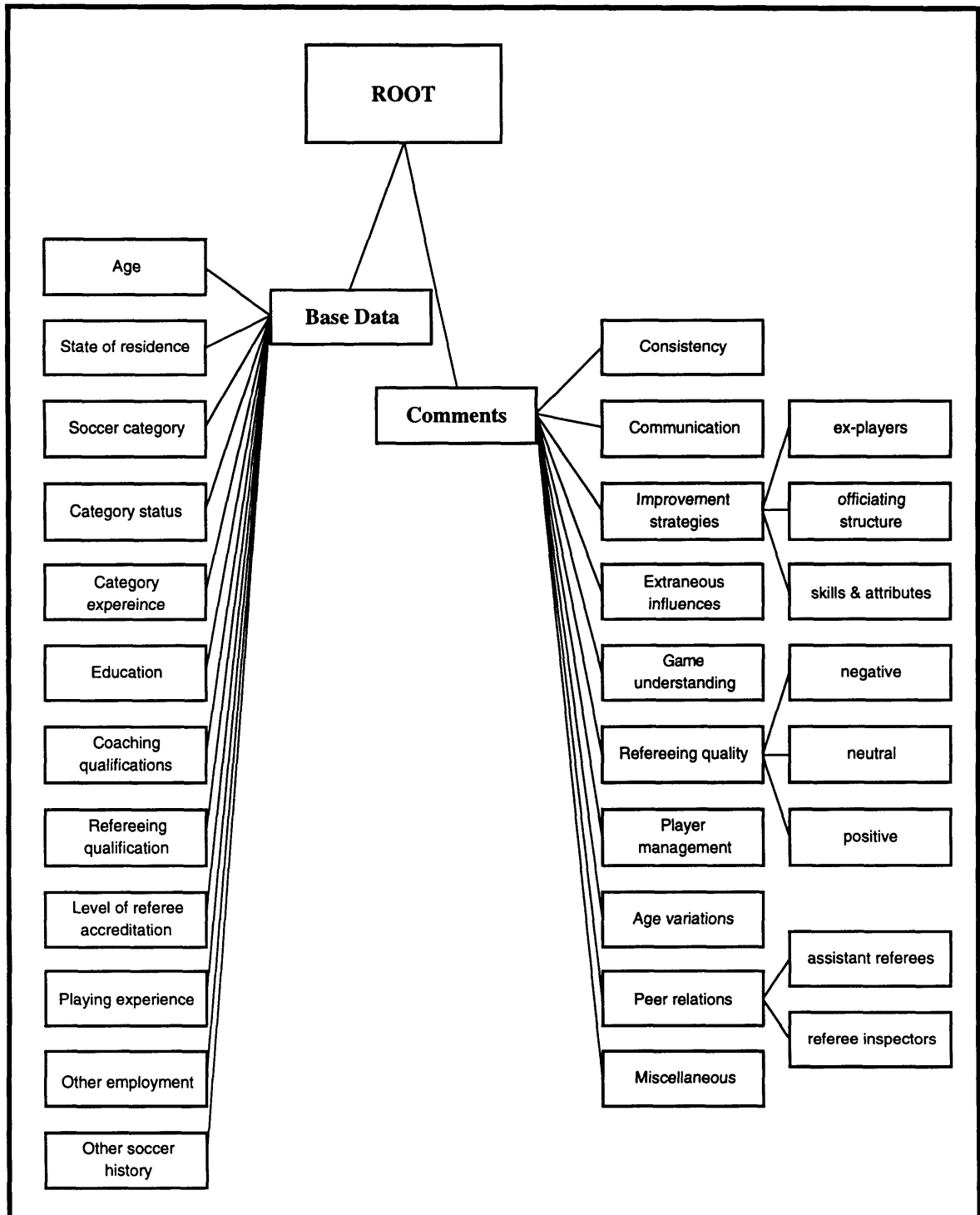


Figure 7.1 NUD-IST tree structure of qualitative data

depth discussion. Examples of the former are provided below¹.

Adapted well to current changes. (108, referee)

No consistency. (111, coach)

Standard reasonably high but too many current referees lack consistency in decision making. (160, referee inspector)

These examples represent the entire response supplied from each of the particular respondents. However, not all responses were this brief. Numerous responses provided comments which contained extensive amounts of information on various refereeing issues. Particularly, responses which mentioned specific refereeing competencies, and made reference to policy and/or training issues that impact on the quality of refereeing, were common. Given the varied information contained in these responses, the comments presented a rich source of data for gaining insight into refereeing performance. An example of such a response, supplied by an assistant referee (respondent 123) is provided below:

My opinion on the current standard of refereeing in the Ericsson Cup is that despite the recent (or perhaps constant, depending on the viewpoint) criticism of Ericsson Cup referees, the standard of most referees is high. It is to be expected that these are in fact the elite referees in this country and should therefore be expected to produce the best performances. However, this does not hide the fact that some improvement is definitely possible and should be pursued.

The one and only thing asked for by most people associated with the game is consistency and uniformity across Australia between referees is one thing that, although difficult, can be achieved with application.

I believe programs for the improvement of all referees are essential and these programs, if conducted on a nationwide scale, can only be for the good of referees everywhere. Refereeing in Australia is obviously improving slightly but hopefully competition with other countries on a refereeing platform (i.e., more Australian referees and assistant referees at FIFA competitions) can only provide extended help to everyone involved.

In summary, detailed comments, such as that above, and the briefer comments detailed previously, begin to disclose the rich complexity which surrounds occupational performance. This complexity is confirmed also through the variety of topics classified under the *Comments* node. In combination, the findings indicate that respondents were motivated to comment on varying aspects of performance, and to differing degrees of specificity.

¹ When presenting examples of respondents comments, all comments are provided as quoted by respondents with two exceptions. Firstly, spelling has been corrected, and, secondly, text is inserted into square brackets to clarify the meaning where necessary. No attempt has been made to alter grammar. Lastly, respondent identification numbers and sub-group membership is provided after each comment cited.

EVALUATIVE IMPRESSIONS OF REFEREEING STANDARDS

Aside from the issues raised in the previous section, comments also revealed assessments of refereeing standards. Within this evaluative context, the following discussion analyses relevant responses. In particular, the distribution and nature of comments which formed a summative impression of the current standard of refereeing is examined. The diversity of opinion between sub-groups is also contrasted. Findings presented in this section address the research question:

Within the context of a free response, how is the current standard of refereeing perceived?

Distribution and Nature of Comments

Comments related to the general standard of refereeing were numerous. Specifically, 175 comments were provided by 90% ($n=107$) of respondents, and were classified into the *Refereeing Quality* node (see Figure 7.1). Comments were classified further into *positive*, *negative* and *neutral* nodes (the titles of these three nodes is intended to reflect the nature of the comments). Typical comments, from within the *positive*, *negative* and *neutral* nodes, are provided below:

Overall the Ericsson cup referees compete favourably with the rest of the world. (152, referee inspector)

The current standard of refereeing in the Ericsson Cup is of a pretty good standard. (130, assistant referee)

I think generally the standard of refereeing in the Ericsson Cup is very good. I believe that most referees work extremely hard to present themselves professionally and to officiate in the fairest possible way. (105, referee)

It's very erratic. (41, player)

Not superb but they are doing fairly well. (12, player)

Standard [of refereeing] this season is average. (162, referee inspector)

The distribution of comments within the *positive*, *negative* and *neutral* nodes is given in Table 7.2. As shown in this table, two perspectives on data interpretation are provided. The first relates to the number of comments classified into the *positive*, *negative* and *neutral* nodes. From this perspective, it can be seen that comments are not spread evenly across the three nodes, with the majority of comments classified into either the *positive* or *negative* nodes. The number of comments classified into these latter two nodes is, in general terms, divided equally (43% and 40%, respectively).

Table 7.2 Summative impression of refereeing standards – distribution of comments classified into the *Refereeing Quality* node

Node	Comments		Respondents	
	N	% of Total	N	% of Total
Positive	76	43	51	47
Negative	70	40	38	36
Neutral	29	17	18	17
Totals	175	100	107	100

The second perspective provided in Table 7.2 centres the number of *respondents* who provided positive, negative and neutral comments. Most respondents (47%) viewed referee performance in a positive context, with the next highest response category (i.e., *negative*) with 36% of respondents. This ‘respondent’ based finding (as opposed to ‘comment’ based finding described previously) provides a stronger indication of agreement concerning the current standard of elite refereeing in Australia. Nevertheless, when all the figures supplied in Table 7.2 are viewed from a broader perspective, it can be concluded that consensual agreement within the cohort, concerning refereeing standards, was not evident.

Group Responses

Evidence to explain the lack of consensus in the overall rating of refereeing performance can be found in comments provided by each stakeholder group. However, before examining responses, it is necessary to clarify how results are presented with respect to the grouping of respondents. In Chapter 6, the five sub-groups which comprised the global sample were collapsed into two larger groups, i.e., the competitive group (players and coaches) and the officiating group (referees, assistant referees, and referee inspectors), for statistical purposes. However, such restraints do not apply always to the analysis of qualitative data and presentation of subsequent results. Therefore, references to specific sub-groups are made from time-to-time, although some findings refer to the competitive and officiating groups when pertinent.

Overall, there was not consensus between the five sub-groups concerning the perceptions of refereeing standards. To clarify the significance of this difference, responses were submitted to the non-parametric Chi-square test. This test is suitable for nominal data where observations can be classified into discrete categories, and treated as frequencies (Burns, 1997). For the test to proceed in the context of this investigation, a contingency table was constructed with comments classified into positive and negative categories, and sorted according to sub-group. The ‘observed’ and ‘expected’ frequencies that accompanied this

analysis were acceptable (expected frequencies should be five or greater in 80% of cells, see Burns, 1997), and a significant difference ($p < .001$, $df=4$) was found.

However, Chi-square does not have the capacity to specify precisely which sub-groups are implicated in causing the significant *alpha* value. To meet this end, inspection of cell frequencies provided a means to interpret the finding. Specifically, analysis of comments classified in the *positive* node showed a majority of comments come from referees, referee inspectors, and assistant referees ($n=40$, 78%)¹.

No referee provided comments in the *negative* node. In fact, all referees who commented on overall standards ($n=11$) indicated a positive attitude to their current performance. However, the referees' officiating colleagues were not as supportive. Of the 23 assistant referees who responded, 12 (52%) offered negative comments, while 7 of 21 (33%) referee inspectors responded with negative views of current refereeing practise. Typical comments from the latter two sub-groups include:

Level between top and bottom [referees] is very large. (121, *assistant referee*)

I feel that the panel B referees are no way near as good as the panel A referees². (125, *assistant referee*)

Some referees who have made the grade appear a little above themselves. (132, *assistant referee*)

... the basic building blocks of knowledge and understanding are well down the league table. (138, *assistant referee*)

There is a gap between the elite referees and 'B' referees in this panel. (154, *referee inspector*)

Of a panel of 14 referees I rate them as follows:

6 – Very competent.

3 – Average competency.

5 – Very ordinary. (158, *referee inspector*)

Notwithstanding the nature of these comments, a general perception emerges from members of the officiating group that the standard of refereeing is satisfactory. In contrast, comments from players and coaches are more pointed, expressed in stronger language, and less

¹ For purposes of clarity, from this point the remainder of discussion utilises statistics relating to frequencies and percentages based on 'respondent numbers.' If statistics are required for 'comments' only, a specific note is made.

² The A and B panels are an unofficial 'grading' of Ericsson Cup referees.

ambiguous to interpret, than those presented by members of the officiating group. Specifically, comments from players and coaches reflect an overall negative impression of refereeing standards. Examples of comments that typified opinion from the players and coaches are provided below:

The standard of refereeing is atrocious. The level is way below what it should be. (55, *player*)

It is not obvious that the refereeing personnel/officials are not up to the required standard. Players commonly ask rhetorical questions such as “is that guy incompetent or a cheat?” Rarely does one complete a game (as a player) and feel the referee has done well. (73, *player*)

The current standard is pretty poor. (86, *player*)

Season 1996-1997 referee’s standard has been below average ... I find it a problem when referees decisions actually costs games. (115, *coach*)

Some positive comments were offered by players and coaches, although most of these came from coaches. Nevertheless, even when positive comments were made, they were usually qualified in some way (see example below).

I’m quite pleased with the overall standard. There have been only a couple of occasions when a referee has affected the outcome of a game through his error. (117, *coach*)

Overall, there was a difference in opinion on the standard of refereeing between sub-groups. Comments offered by the players and coaches represent a clear dissatisfaction with the overall standard of refereeing, while referees, assistant referees and referee inspectors view the current standard more positively. The lack of general consensus between sub-groups can possibly be explained by the differing roles each group assume in the soccer industry.

Discussion and Implications

Results presented in this section provide additional insight into the standard of elite refereeing in Australia. From the perspective of Research Question 5.1: *Within the context of a free response, how is the current standard of refereeing perceived?* it is apparent that the performance of referees is perceived to be variable.

One reason to explain some of the diversity in performance appraisal relates to occasional comments referring to ‘A’ and ‘B’ panels. These are unofficial lists, developed by Soccer Australia, which classify Ericsson referees according to ability. The top referees are listed in the A panel, and the next tier of referees comprise the B panel. Comments indicated that performance variation between these groups is substantial. This differentiation has arisen despite referees from both ‘panels’ refereeing within the same level of competition. Inasmuch, the finding implicates a marked performance ‘gap’ within the ranks of elite referees.

Given the small number of referees recruited each year to officiate elite soccer, the finding of a performance ‘gap’ is noteworthy. Specifically, two implications arise. Firstly, Soccer Australia may need to review and/or refine its recruitment strategies to ensure that only the best referees are selected for Ericsson Cup duties. The current practice of recruiting referees from all states has the potential to marginalise capable referees from the more populous states. Alternatively, if the best possible referees are being recruited, they are not overcoming perceived performance deficiencies. This latter scenario, if substantiated, provides evidence for referee training and development strategies to be altered to reduce the performance ‘gap.’

However, the differences in performance standards, across the referee cohort, do not explain the divergent opinions between sub-groups concerning the overall quality of elite soccer refereeing. Specifically, referees, assistant referees, and referee inspectors perceive the performance standards of referees favourably (although this feeling was stronger among the referees than the latter two sub-groups). In contrast, players and coaches express dissatisfaction with current performance levels. Given the assumption that all sub-groups view the same refereeing performance (and all have a relatively homogeneous *knowledge* of soccer), the differing soccer role each sub-group assumes would seem most likely to account for the varying opinion.

Particularly, players and coaches are most preoccupied with the outcome of a game. This being the case, players and coaches are more likely to base referee assessments according to perceived ‘referee facilitation’ in achieving game outcomes. For example, a coach or player may assess a referee’s performance based on one critical decision that resulted in a goal. However, in the context of the whole game, the decision may not be reflective of the entire refereeing performance. Alternatively, referees, referee inspectors and assistant referees have little concern for which team wins or loses. They are required to focus on soccer Law administration, and, in this, their role is to focus on a multitude of decisions across a variety of contexts. Consequently, a broader perspective is possibly bought to the evaluation by this group.

Three other reasons are also proposed to explain group differences. Firstly, it is hypothesised that when assessing refereeing performance, the sub-groups value competencies differently. If the validity of this hypothesis is accepted, the divergent values ‘go to the core’ of the present study. Given the scarcity of clearly defined refereeing performance criteria, it is not surprising a discrepancy in opinion exists. Lack of clarity and understanding, in any form of performance appraisal, invites divergent and contradictory opinion. This, in turn, is compounded by a lack of agreement about what are, and what are not, important refereeing competencies (this finding is consistent with ‘competitive’ and ‘officiating’ group differences reported in Chapter 6, and is discussed later in this chapter).

Secondly, it is probable that referees (and to a lesser extent, assistant referees and referee inspectors) do not view refereeing performance with the same 'critical eye' as players and coaches. This is based on the perception that, in general, people view performance in more favourable terms if they have some degree of professional or personal affinity with the person/s under review. This phenomena, analogous to a 'halo effect,' is likely to be exaggerated in the assessment of refereeing standards, primarily due to the 'them versus us' mentality that prevails between members of the officiating and competitive groups.

Lastly, sub-group differences can also be explained in terms of proximity to the referee and the field of play. Coaches, assistant referees and, to a certain extent, players, are removed from specific game incidents by varying distances. However, full possession of circumstantial information is fundamental for any valid assessment. For example, the quality of referee communication via hand signals can be assessed from positions well away from the immediate area of play (in reference to this example, it could be argued that the better the competency is performed, the further away the assessor can be, i.e., a referee inspector in the grand-stand). Alternatively, assessment of many other refereeing decisions, e.g., over-the-ball tackles, hand-balls, etc., are more valid when the assessor is close to the play. As this issue of 'proximity to play' is variable with respect to each group of soccer stakeholders, conflicting opinion among sub-groups concerning refereeing competence is perhaps unavoidable.

In conclusion, two points crystallise the discussion in this section. Firstly, the diversity of opinion described in the results and discussion is likely to be symptomatic of referee performance assessment across other sports. This would be evident especially in sports that are inherently similar to soccer. For example, basketball, netball, and other football codes have a combative nature, and present ambiguous game situations framed by standardised rulings. Moreover, for a particular decision given in favour of one team, the other team (and respective coaching staff) is going to perceive they are on 'the wrong end' of the decision. This circumstance has the potential to compromise the objectiveness of subsequent refereeing assessments, irrespective of the validity of bias perceptions.

Secondly, discussion has illustrated the need for general agreement concerning how effective refereeing performance is judged. Moreover, agreement needs to be achieved with all stakeholders having ownership of the criteria on which judgements are based. Without criteria developed in this manner, disagreement between sub-groups is likely to continue.

SPECIFIC COMMENTS ABOUT REFEREEING PERFORMANCE

The previous section gave general evaluative impressions of refereeing standards. However, it did not address comments related to specific aspects of performance, nor peripheral factors which impact on the quality of referee performance. Accordingly, discussion in this section examines such comments, and answers the research question:

Which aspects of refereeing performance were the focus of most comment?

Overview of Specific Comments

Comments from respondents encompassed an array of refereeing competencies and issues. This diversity was flagged initially in the NUD•IST tree structure (see the ‘children’ of *Comments* node). More specifically, Table 7.3 details the frequency of responses with respect to node and sub-group. At face value, the table reflects the assorted topics at the fore-front of refereeing performance, and the emphasis placed on each topic by each sub-group. Moreover, figures contained in Table 7.3 allow strengths and weakness of elite soccer referees to be identified, and potential improvement strategies to be explored.

Table 7.3 Content analysis – node frequency distribution by group

Node	Referees % n=11	Assistant Referees % n=24	Referee Inspectors % n=23	Players % n=42	Coaches % n=18
consistency	27 (3)	21 (5)	13 (3)	36 (15)	39 (7)
communication	9 (1)	8 (2)	13 (3)	48 (20)	17 (3)
extraneous influences	36 (4)	17 (4)	22 (5)	19 (8)	17 (3)
improvement strategies	27 (3)	71 (17)	65 (15)	48 (20)	17 (3)
game understanding	0 (0)	4 (1)	4 (1)	29 (12)	11 (2)
player-management	0 (0)	25 (6)	22 (5)	31 (13)	6 (1)
peer relations	9 (1)	13 (3)	17 (4)	5 (2)	6 (1)
age variations	18 (2)	42 (10)	43 (10)	12 (5)	11 (2)

With respect to the proportion of responses provided in Table 7.3, uniformity among sub-groups is variable for some nodes. For example, in the *communication* node, 9% ($n=1$) of referees made comment, as opposed to 48% ($n=20$) of players. Other nodes which showed notable differences between two or more sub-groups were *player-management* and *game understanding*. Nevertheless, other nodes did present response-rates that were relatively consistent across sub-groups, i.e., *consistency*, *extraneous influences*, *improvement strategies*, and *peer relations*.

Of these four nodes, *improvement strategies* attracted the most comments. However, the figures quoted in Table 7.3 require clarification. In particular, some comments in the *improvement strategies* node were classified into a second node also. This is illustrated in the following example, where the comment was coded in the *improvement strategies* and *player-management* nodes (the comment relates to how a referee interacts with players, and has ramifications also for how referees can improve performance):

The type of players are changing and insight into how to calm players, defuse situations, understand the difference between frustration and anger, for example, is extremely important. (157, referee inspector)

Although the cross-referencing of comments, such as that offered by respondent 157, has the potential to confound the frequency distribution of comments to some degree, the procedure should not be seen as a restriction on the data interpretation. Indeed, the classification of comments in this manner reflects the complexity in describing occupational performance.

Due to the ‘dual classification’ of some comments, the *improvement strategies* node is not examined as a single classification category. Instead, comments from the node are interwoven with discussion on other topics of note, including consistency, player-management, age variations of referees, and the influence of extraneous issues on the refereeing process. To aid clarity, discussion is separated into two sub-sections. The first examines comments concerned with specific refereeing competencies, and the second sub-section explores responses related to issues that are removed from on-field performance.

Refereeing Competencies

Of the performance criteria listed in Table 7.3, the comments mentioned most frequently relate to *consistency*, *player-management*, and *communication*. These are examined at length in the following discussing, with particular focus on the differing perspectives provided by various sub-groups.

Consistency

Due to the strong emphasis on the decision-making role of officials, consistent decision-making is seen as a hallmark of good officiating (see Chapter 1). The nature of some sports facilitate consistency through rules that require unambiguous decisions (e.g., in tennis, the ball is either ‘in’ or ‘out’). However, in the highly contextualised environments of soccer matches, consistency is more problematic.

Nevertheless, despite the difficulty in applying soccer rules, consistent decision-making is highly desirable. The importance of consistency is reflected in the quantity of comments offered by respondents, where, as a single area of refereeing *performance*, this competency generated most comment (see Table 7.3). In particular, 36% of players and 39% of coaches

commented on consistency. All players and coaches who made comments wanted improvement in this performance criterion. As stated by one player:

Referees have been giving ambiguous decisions and consequently costing teams vital points. (55, *player*)

The concerns expressed relate primarily to three factors, namely, consistency of decisions within a game, consistency of decisions from game-to-game by the same referee, and consistency of decisions between different referees. Examples of comments related to each of these forms of consistency are, respectively:

Lack of consistency by the referees in handing out fouls throughout the match. Similar fouls early in the game are not dealt the same as later in the game. If referees give a decision he/she has to reward the other team in the same way if the same or similar foul occurs. (38, *player*)

I have seen many referees perform well in some games, then when he referees your game, you think, he's not too bad a referee, but then he has a terrible game. (66, *player*)

Overall there is a large inconsistency in the refereeing ranks. Players and coaches know that a certain team will get away with one thing with a referee compared to what they can get away with another referee. (173, *coach*)

Consistency issues were also raised by members of the officiating group. However, these comments were less numerous (18%, $n=11$), and are more variable than those offered by the competitive group. While a minority of comments were similar to the opinions expressed by players and coaches, the majority expressed a more positive perspective. Additionally, opinion tended to take account of the complex role of the referee, and, in one instance, questioned the validity of coaches to question the consistency of referee decision-making. As such, these comments represent a contrast to those expressed by players and coaches. Examples of comments provided by the officiating group include:

I think the standard overall is pretty good allowing for the inconsistencies with different styles. (102, *referee*)

Each referee have their own personality and each soccer game is never the same so with this in mind one will always have inconsistency. (134, *referee assistant*)

Define consistency? A referee will view a incident from a different view than a coach or spectator, yet they whinge about consistency. (99, *referee*)

It is evident that consistency issues present a major point of conjecture across, and, in some instances, within groups. A possible reason for this is the ambiguous and contextual nature of referee decision-making. Many factors, including player skill levels, climatic conditions, and referee personality, can present non-constant variables. Factors, such as these, complicate the refereeing role and the subsequent assessment of refereeing performance.

Nevertheless, it is clear that consistency, as an indicator of effective refereeing performance, requires clarification (i.e., what are consistency expectations?). To some extent, this is addressed in comments that provide practical examples of refereeing inconsistencies. Predominantly, the consistent awarding of free kicks and sanctions is most common. Within this example, perceived favourable treatment for some categories of players is prevalent. The following comments typify responses related to this example:

[the referee] appears to referee individual players according to their reputations and not on specific incidents. For e.g., flamboyant players seemingly always pulled up for diving where upon fouls have actually been committed. Similarly, players with 'clean' images get away with 2 footed tackles. (73, *player*)

Double standards for 'well known' players and rookies. (43, *player*)

Australian [i.e., international] players in Ericsson Cup seem to get special treatment when committing fouls and no cards eg [an ex-international] will foul 9-10 times a game and not get a yellow card. Refs must be to scared to caution him and his mates. (171, *assistant referee*)

Compounding such concerns are inconsistent interpretations for the same or similar incidents. For example, it is evident that players and coaches expect that the same sanction be applied to the same foul, no matter what contextual or mitigating circumstances surround the incident. This is evidenced through comments provided below:

Sometimes the interpretation of tackles is different to the opposing team. (7, *player*)

They [the referees] have difficulty in distinguishing between deliberate and accidental fouls. (62, *player*)

In summary, comments provided in this discussion were pointed and passionate in many instances. The emotive nature of responses can be explained by a number of factors which are likely to complicate perceptions of consistency competence. These factors include: the variable contexts of refereeing decision-making; related equity and 'natural justice' issues that surround refereeing decisions; and, the multi-dimensional characteristic of consistency. Nevertheless, though the provision of specific game-examples where inconsistencies commonly occur, comments provide pointers for minimising the occurrence of inconsistent decision-making, irrespective of the game context.

Communication and Player Management

In Chapter 1, discussion noted that referees are required to communicate decisions through verbal and non-verbal means. Moreover, discussion recognised that referee-player interaction can occur in volatile playing environments (e.g., players are highly motivated and not always agreeable to refereeing decisions). Accordingly, the manner in which referees

relay decisions, and player reaction to refereeing decisions, can be extreme. These circumstances help to explain the high proportion of comments attributed to communication and player-management as a single area of referee performance (see Table 7.3).

The highest proportion (47%) of communication comments came from the players. The quantity of comments from this sub-group can be explained by players engaging in *on-field* communication, with the referee, more regularly than other sub-groups. Although player comments vary, they indicate a need for improved rapport between referees and players. Examples of such comments include:

If players and referees can develop a friendly relationship I believe it could help the game. (58, *player*)

We [players] are human beings, they can still talk to us as normal people on the pitch. (7, *player*)

To whoever is listening ... do you realise how disheartening it is to hear that referees are reprimanded because they engage in bridge building with players i.e. they dare to establish relationships and rapport. (73, *player*)

Most generally, comments provide evidence that referees avoid communicating with players over-and-above the basic requirements of relaying decisions via whistling and hand-signalling. This finding is not remarkable as the Laws of Soccer do not require referees to explain rulings to players (similarly, team captains have no special rights to question referee decisions). However, a strong desire for more personalised communication came from players, coaches, referee inspectors, and referee assistants (no referees made comments on this topic). A representative sample of pertinent comments are provided below:

More communication is needed between players/officials to understand the situations that may arise. (12, *player*)

Their [the referees'] communication with players is very low. Most of them swear at players. This is not on. If a player swears at a referee or official they will get a yellow or red card. It must go both ways. (42, *player*).

They need to create a more friendly atmosphere amongst players. More person to person talking with the players. (119, *coach*)

Quite often when a red or yellow card is shown, it is almost encouraging a verbal backlash, as a number of referees, either flick the card or just about stick it in the offenders face. The type of players are changing and insight into how to calm players, defuse situations, understand the difference between frustration and anger, for example, is extremely important. (157, *referee inspector*)

With the ever changing role of the referee ... being able to communicate more effectively, and acceptance that you can not always deal with individuals the same way should ensure harmony and good will among players, officials and the general public, are more necessary than ever. (127, *assistant referee*)

Further analysis of comments such as these revealed three issues which expose underlying causes of player-management/communication problems. These are: the age of the referee; referee compliance to strict rule interpretation; and, the manner of referee communication with players. Of these three issues, the age of the referee drew the greatest amount of comment (player comments tended to predominate, although a cross-section of the sample, i.e., referee assistants and referee inspectors, offered similar opinion). Most particularly, comments raised doubts about the capacity of younger referees to have appropriate player-management strategies, e.g.,

[young referees are] Inconsistent, inexperienced, unapproachable and generally of poor standard. (50, *player*)

It is difficult for many players to respect a referee who is 22-26 years of age because it is evident that at such a young age, he lacks experience and more often than not has difficulty establishing a rapport with older experienced players. (139, *referee assistant*)

Communication with and attitude to players needs attention. This is probably because many Ericsson Cup referees are reasonably young (early to mid twenties) and have not yet had the experience of 'man management' – many still tend to take a school teacher attitude by lecturing players in a semi-arrogant attitude and do not get the respect of players, coaches or club officials. (160, *referee inspector*)

This issue of referee age was apparent also in comments relating to the strict application and adherence of rules. Comments from the players, assistant referees, and referee inspectors, reflect a desire for referees to be less stringent in their rulings and dealings with players. It is evident that a more 'relaxed' attitude by the referee may improve player-referee relationships. Typical comments supporting this assertion include:

You find that some of our referees have a good rapport with players on the field where others, I find, are to rule stringent. It is black and white to most of them. No talking at all. (66, *player*)

With the younger generation of referees, they possess high levels of fitness, knowledge of laws and technical ability but need to increase their man management skills. (140, *assistant referee*)

Today there is little DISCRETION with discipline, whereas years past, referees had a lot of discretionary power. The interpretation of the power caused referees to be more mature and experienced to control players. (151, *referee inspector*)

Despite the overwhelming call for referees to be more responsive to meaningful communication, comments suggest an acceptance for communication to be a two-way process. Players, in particular, recognised that communication with referees needed to be predicated on mutual respect between groups, for example:

The referees that I consider competent are those who effectively communicate with players, with equal respect; and if you [the referee]

speak to a player in an even or friendly tone, I think you'd find players will respond in kind. (9, *player*)

Overall, the thrust of comments highlighted a fundamental concern with the communication skills of elite referees. Particularly, the soccer stakeholders expressed concern with the manner in which referees communicate with players. Given the central arbitration role of the referee, and the emotive environment in which elite referees work, communication will continue to be a prevalent topic of debate. Nevertheless, referees need to be cognisant of their on-field actions and speech, and how these inter-personal factors impact on related player-management issues.

Refereeing Issues

Findings thus far have highlighted concerns relating to a variety of refereeing competencies. However, as noted in Table 7.3, data were coded also into four of nodes which are 'issue' based, rather than 'competency' based. These are *improvement strategies*, *extraneous influences*, *game-understanding*, and *age variations*. These data provide opinion on a variety of refereeing issues that impact on the overall quality of refereeing performance. Two themes connect the data, namely, referee recruitment strategies of Soccer Australia, and the inspection process used to assess referee performance.

Referee Recruitment

Recruiting people to undertake officiating duties is a major problem for many sports. As noted in Chapter 1, the recruitment and retention of referees is a perennial and complicated issue. However, if sporting organisations are to maintain a viable officiating cohort, and improve officiating standards, finding a solution to the problem is essential.

Numerous respondents (37%, $n=44$), from all sub-groups except referees, provided comments on this issue. Discussion tended to take one of two paths, i.e., the strategy of Soccer Australia to promote actively younger referees (usually in their mid-twenties), and the need to attract former players into the refereeing ranks.

The promotion of younger referees generated considerable discussion. Overall, there were differences in opinion among sub-groups concerning the strategy's success. Examples illustrating the diversity of comments are provided below:

I feel that the transition from young referees to experience has been gradual and highly successful. (127, *assistant referee*)

The younger the referee, it seems, the greater the motivation to adapt to changes ... They are more willing to accept constructive criticism [than older referees] and attempt to make use of it to improve their skills. (148, *referee inspector*)

I think that some of the referees are inexperienced and don't try to adjust to the level of the game. (64, player)

Basically the younger refs have a lot to learn. (94, player)

Despite a lack of consensus, agreement was found on three aspects of the youth promotion strategy. Firstly, the fitness standards of the younger referees was appreciated and rated highly¹. Comments such as the one below reinforce the value of young referees to the maintenance of this important performance criteria:

The fitness of most referees has improved considerably over the last few years and with the policy of promoting youth the current referees are probably the fittest. (137, assistant referee)

Secondly, a general perception exists that young referees have the capacity for improvement, and, as such, can become world-class if given appropriate time and training. This was evidenced from the type of comment listed below:

With some of our younger referees the potential for them to achieve higher recognition looks very rosy. (129, assistant referee)

The implication of these first two points is that, in the long term, there are advantages in persevering with the promotion of younger referees. However, the third aspect of consensus brought a counter-balance to this positive view. Respondents, across all sub-groups, noted a lack of communication and player-management skills shown by the younger referees. Discussion of this deficiency, with respect to younger referees – and referees in general – has been undertaken previously. It is suffice to note that, overall, the lack of proficiency in the execution of player-management and communication skills are seen to be the greatest failing of the youth program, and data suggest training programs should focus particularly on the development of these two skills.

Comments did present, however, a strategy which has the potential to circumvent the communication/player-management concerns, i.e., the recruitment of ex-players into refereeing. A context for this strategy is that, players, generally, tend to stop playing elite soccer when they reach their late twenties or early-mid thirties. The higher age bracket of ex-players (compared to younger referees), coupled with their experiences in dealing with other elite soccer stakeholders (normally over a number of years), has the potential for a more considered approach to be brought to communication/player-management needs. Given these circumstances, the recruitment of ex-players into elite refereeing can aid in negating 'immaturity' concerns associated with the younger generation of referees. Supporting comments include:

¹ Fitness was seen as a major strength of all referees. This issue is discussed in detail in the final section of this chapter.

If possible in the future, players from an elite level should be swayed into refereeing when their careers are over. This would earn more respect from players. (6, *player*)

Past players should be encouraged to referee as they understand the pressures etc of the game. (70, *player*)

More specifically, comments, such as those of respondents 6 and 70, imply a desire for referees to be more empathetic towards the attitudes, emotions, and motivations which drive player behaviour and actions. Such understandings, or lack thereof, is the target of direct criticism of current referees (of all ages), particularly from the players. The following comments typify responses:

Should have actually played the game somewhere along the line. (54, *player*)

They should have an understanding of the game and not just the rule book. (50, *player*)

They need more experience in playing the game to understand why they make decisions. They need to have played or have more experience at playing the game at a higher level to understand how players think. (62, *player*)

To have played the game is a definite advantage. (9, *player*)

Players always feel that had a referee played soccer at a reasonable level then it would benefit them on the field. It would help them understand the passions, pressures and desires of players during the 90 minutes. (66, *player*)

The validity of these comments is given some degree of verification when the playing history of referees is examined. From the referee cohort used in the present study, over half ($n=9$) listed 'local associating soccer' (i.e., amateur soccer) as the highest level in which they had played. Of these, three had played at the junior level only. No referee had played the equivalent of Ericsson Cup soccer, although three had played soccer at the state league level. This evidence provides a reason why referees, in general, are perceived to lack an understanding of player motivations and actions in tense game situations at the elite level. The recruitment of former Ericsson Cup players has the potential to avoid such perceptions.

In a related recruitment issue, attracting ex-players into refereeing careers is problematical given refereeing remains a part-time pursuit. This is despite the general move towards full-time professionalism for other stakeholders, i.e., players and coaches. Given this circumstance, for ex-players to be attracted into elite refereeing (or, for that matter, anyone else of potential ability) comments suggest a financially viable career-path needs to be established. For example:

I believe to improve the standard of refereeing, we must ... be paid more for the highest level in your country. (130, *assistant referee*)

The [standard of] referees leaves a lot to be desired due to, in my opinion, that the current referees are required to maintain full time employment, hence restricting the amount of time available for training. (133, assistant referee)

In my opinion the only way to improve on what we have is to put as many hours into training as the player and coaches that we are expected to control do ie possible full-time refereeing. (141, assistant referee)

... high profile refereeing is not a hobby, but a full time commitment to excel in top level sport. (156, referee inspector)

Reasons for the part-time status of referees are numerous, but none more prevalent than the availability of money to fund such a venture (Power, G., Director of Referees for Soccer Australia, 1999, pers. comm., 30 June). Nevertheless, comments indicated referees would be more likely to commit themselves more to their role if restraints on training times that exist currently (due to the part-time status of elite refereeing) could be removed. This would increase the likelihood of a more thorough and professional approach to addressing technical weaknesses, and building upon performance strengths (in similar ways players do through their training).

The issues raised through comments in this discussion illustrate the interaction of factors that underlie referee recruitment at the elite level. Evidence suggested the balance between youth and experience, and the respective advantages that each group presents, are central to strategic recruitment. Nevertheless, whatever the 'formula' used to achieve this youth/experience balance, improved recruitment approaches are pivotal to improved refereeing standards.

Referee Inspection Process

The final aspect of elite refereeing which drew marked comment was the contribution of referee inspectors to referee assessment. As detailed in Chapter 4, all refereeing performances in Ericsson Cup games are assessed by a referee inspector appointed by Soccer Australia. The inspector provides a 'quantitative' mark based on criteria developed by FIFA, plus written comments based on the inspector's perception of the referee's performance. This assessment forms the basis of future game selection, and provides direction for referee coaching.

The value of inspector assessments is given various perspectives by several sub-groups in the written responses. Referees and assistant referees (37%, $n=13$) provided the greatest amount of comment, followed by inspectors (17%, $n=4$). Only one coach and one player provided comment.

The pattern of responses can be explained by two inter-related reasons. Firstly, referees and assistant referees have the most to 'fear' from inspector reports, i.e., inspector assessments are used as a basis of referee/assistant referee appointments. Therefore, it is probable that referees and referee assistants welcome the opportunity to comment on their 'judges.' Secondly, players and coaches have no direct professional relationship with referee inspectors. Thus, members of the competitive group are probably less likely to provide comments on referee inspectors as the actions of inspectors have minimal impact of the work of players and coaches.

Specifically, the majority of comments from referees and assistant referees were not complimentary. Comments tended to focus on inspectors' lack of refereeing experience at the elite level evident. Examples of these comments include:

A contributing factor to the standard of refereeing, is the standard of inspectors. Many inspectors have not officiated at a very high level of soccer and only a few have actually refereed on the National League. (139, *assistant referee*)

At the moment there are no inspectors inspecting referees at the top level that have refereed at the top level. (130, *assistant referee*)

Aside from the refereeing experience of the inspector, a secondary concern surrounding inspections is the criteria on which inspectors assess referees. It was claimed that:

Competencies are 'measured' inadequately ... So the messages sent back to referees after each performance are inconsistent, disorganised and open to question. As a consequence refereeing is similarly inconsistent and disorganised. If the assessment of competencies through on-field assessment is bad then performances can be random. (138, *assistant referee*)

Statements, such as this, raise the concern that assessment criteria lack validity, with resulting feedback lacking relevance. This is of some concern, as improvement strategies predicated on mis-informed feedback are unlikely to lead to improved refereeing standards. However, it should be recognised that invalid criteria is not a direct fault of the referee inspector, but impacts, nonetheless, on the effectiveness of the inspector's role.

Lastly, two additional aspects of the inspection process also warrant mention. The first relates to the use (rather than the validity) of assessment criteria. It is evident that the inadequacy of assessment criteria is compounded by inspectors who are not *au fait* with the use of assessment instruments. Secondly, statements indicate that the integrity of the inspection process maybe doubtful. The following comments support these two assertions:

Referees inspectors are not being taught effectively how to use evaluation reports to enhance competency and professionalism. (164, *referee inspector*)

Inspectors appear to be subjective rather than objective and are inconsistent when appraising the performance of different referees. (139, *assistant referee*)

The referees have openly said they referee according to the state they are in, and also the inspector they have. (157, *referee inspector*)

With the so called top refs getting this special friendly treatment and not been made an example of for poor performance they have got complacent and the standard has dropped. (171, *assistant referee*)

Given the many concerns with the inspection process illustrated in this discussion, the role of the inspector, in improving refereeing standards, is not being maximised. Ideally, adequately qualified and trained inspectors, using valid and unambiguous performance criteria, should play a fundamental role in improving refereeing standards. This is particularly the case through their role in the 'feedback loop' for refereeing training and education. Through accurate performance appraisal, quicker diagnosis of refereeing problems is possible, thus allowing specific and immediate remediation strategies to be implemented. In doing so, the potential exists for the 'troughs and peaks' in refereeing performance to be minimised, thereby producing more consistent performances.

Discussion and Implications

Comments and issues addressed in this section have confirmed the complexity and intricacies of refereeing performance. Obviously, the production of better referees requires a multi-faceted approach, and warrants the consideration of numerous impacting variables. Accordingly, in response to research Question 5.2: *Which aspects of refereeing performance were the focus of most comment?* a range of performance aspects are relevant. These included specific refereeing competencies, and officiating issues that relate to the overall quality of refereeing performance.

With respect to individual refereeing competencies, two points emerge. Firstly, the consistent application of rules drew considerable comment. Clearly, players and coaches view variation in the application of the Laws of Soccer unfavourably. They expect these laws to be applied consistently by the referee. This expectation applies regardless of who the player is, in what position he plays, or the actions he undertakes. One example cited by respondents is when certain categories of players are given preferential treatment (for example, international players are perceived to receive favourable decisions due to their high status). As examples such as this are noted explicitly, inconsistent decision-making circumstances can be brought to the attention of referees. With referees more cognisant of their actions, the possibility of improved consistency is not only likely, but achievable.

In a related point, it should be noted that the Laws of Soccer provide referees with the capacity to minimise variation in rule application. For examples, specific sanctions are

mandated for specific fouls. Improvements in consistency rely on referees being mindful of these sanctions, and applying them to minimise variability in decision-making whenever possible.

Nonetheless, comments indicated the necessity of a thorough examination of the concept of consistency, and how it is applied. Implicit in this undertaking is one central notion, namely, What are realistic consistency expectations? Given the complexity in clarifying consistency, attempts to answer this question requires substantial consideration. However, any subsequent process must embrace a wide cross-section of soccer stakeholders. In particular, referees and players should be central to this process given the direct responsibility of the referee for making game-rulings and players bearing the direct consequence of refereeing decisions.

The second main point to emerge from comments concerned *communication* and *player-management* issues. The comments provided suggest that these competencies also require improvement, and this is needed on two fronts. The first is that referees should talk to players and explain the reasons for decisions that are made. By providing players with reasons for refereeing decisions, there is an increased probability of players being more accepting of decisions. Although the letter-of-the-law provides referees with an ‘out’ to avoid direct verbal communication with players, comments made by stakeholders were emphatic that any ‘extra’ communication offered by the referee has the potential to improve referee-player relations.

Secondly, the way messages are conveyed by the referee necessitates attention. Comments from players and officials indicated referees lack competence in this regard. Examples of referees taking a ‘holier than thou’ attitude, and being ‘in the face’ of players were evident. Improvements in this performance area are clearly of benefit to the referee by inducing less player confrontation through their actions. Given that player confrontation is cited consistently in the literature as a major cause of stress and referee drop-out (see Chapter 1), a less confrontational environment may improve the likelihood of referees enjoying their work, and encourage them to stay in the profession longer.

Beside these competency-based issues, other issues arose that also have an impact on refereeing standards. The first is the recruitment of ex-players into refereeing. Former players bring with them vast amounts of playing experience, and possibly possess greater insight (than current referees) into the reasons players behave as they do. Specifically, ex-players are more likely to have an understanding of player motivations and the pressures under which players perform. Understandings such as these can provide a unique perspective on the refereeing process.

Nevertheless, the recruitment of ex-Ericsson Cup players into elite refereeing has been problematic. Part of the difficulty in recruiting this cohort lies in the financial attraction of a profession that is less rewarding than playing. Currently, referees earn a fraction of player salaries (this is despite refereeing duties consuming similar amounts of time and physical effort). Until this fiscal discrepancy is addressed, the attraction of former players to the ranks of refereeing is likely to be low.

Associated with monetary remuneration and the recruitment of ex-players is the notion of a viable career path. The provision of long-term advancement and promotion, and attractive levels of reimbursement, are fundamental to any occupation trying to attract quality employees. In this regard, elite soccer refereeing is restrictive. Soccer Australia has an age ceiling, in line with FIFA policy, on all elite referees, i.e., referees must be younger than 45 (the assumption being that once this age is reached, required fitness levels become problematic). Accordingly, players may see this limit as prohibitive, particularly if they retire from playing in their mid-thirties (such retirement ages are not uncommon). The implications here are two-fold. Firstly, Soccer Australia could remove the age restriction on Ericsson Cup referees, and tie eligibility to *ability* and *fitness*. Secondly, if age restrictions remain, Soccer Australia may need to offer accelerated progression to ex-players. This could mean having ex-players refereeing elite soccer, if ability permits, within 3-5 years (maximum) of playing retirement. This provision allows for a potential career in elite refereeing of at least five years (form permitting).

Notwithstanding these monetary and career path concerns, the recruitment of younger 'career' referees by Soccer Australia is seen as a strategy of merit. It was argued that the promotion of younger referees would establish higher standards of refereeing in the long-term. Particularly, their technical ability and fitness levels have drawn high levels of praise by soccer stakeholders. However, an obvious deficiency has emerged in their performance, i.e., player-management and communication skills. It is clear that referee training programs for this cohort need to provide a substantial focus on these refereeing skills. A logical consequence of improvement in these skills is better rapport with players, which, in turn, makes refereeing and playing more enjoyable and satisfying.

Given the implications of ex-player and youth recruitment strategies, it would be logical for both policies to be seen as complementary, and promoted in tandem. The combined strengths, which each strategy brings to the refereeing role, counters perceived weaknesses within each strategy. Moreover, given each strategy is pursued with equal rigour, the over-representation of referees from one recruitment base could be minimised. As a result, the chances of the same refereeing deficiencies emerging on a regular basis would be minimised. More importantly, the requisite skills which each group brings to the refereeing process would diversify the peer teaching processes. Former players afford a unique perspective to

career referees, while the career referees would provide invaluable technical advice to ex-players. Again, the complementary benefits of cross-recruitment has potential to improve the overall standards of refereeing performance.

Whatever the benefits of diversified referee recruitment bring to elite soccer, these benefits will be marginalised unless the referee inspection process is modified. Comments provided substantial evidence that the inspection process requires review. This is necessary from two perspectives. Firstly, assessment criteria need to be more explicit. Comments noted that the performance criteria were too vague, thus increasing the subjectivity of assessments. Secondly, and compounding the first, is the legitimacy of inspectors. Many inspectors have not refereed at a standard similar, or close to, Ericsson Cup competition. Comments were unequivocal that inspectors need to possess an understanding of elite refereeing in order for their assessments to have legitimacy.

In summary, findings presented in this section reflected fundamental concerns about the performance of elite referees. Moreover, findings indicated that improvements in refereeing performance are not related necessarily to on-field refereeing performance, but can be traced to the ‘infrastructure’ surrounding the refereeing process.

INTEGRATION OF WRITTEN COMMENTS AND QUANTITATIVE RANKINGS

Results presented in the two preceding sections have provided insight into the performance of elite soccer referees. In doing so, evidence has emerged which articulates and augments the results from the Likert scale responses presented in Chapter 6. The links are elaborated upon in this section, particularly in reference to the assessment of refereeing standards, and the performance of specific competencies. Findings presented in this section address the research question:

How do qualitative comments substantiate, and elaborate upon, the quantitative findings?

When articulating the written comments with the quantitative rankings, links are examined in the context of both *importance* and *preparedness* rankings (although an emphasis on *preparedness* ranks predominate). This approach facilitates examination of competencies that are perceived to be performed with low and high levels of *preparedness*, plus it helps to explain the differing opinions concerning refereeing standards described earlier in this chapter.

Competencies Performed with Low Levels of Preparedness

As detailed in the second section of this chapter, the performance criteria of consistency, player-management, and communication, attracted considerable comment (see Table 7.3). The comments were predominantly negative, and came from all sub-groups except referees. Inspection of the competency rankings, with respect to *preparedness*, helps to explain this finding (see Table 7.4). Specifically, rankings for competencies aligned to consistency, player-management, and communication were low. As noted from Table 7.4, the pattern of low ranks for these three competencies were consistent across the competitive and officiating groups.

Table 7.4 Low ranked competencies – Preparedness

Issue	Competency	Preparedness Rankings		
		Competitive Group	Officiating Group	Overall
Consistency	2.1 applies the Laws of Soccer consistently (within each game and over the season)	36	26	34
Player-management	2.8 manages conflict (communication with players, use of presence and personality)	35	29	34
Communication	3.4 communicates (verbal and non-verbal) with players on and off the field	34	30	33

The written comments which support these rankings have been discussed at length previously, and are not repeated in detail here. However, of interest is the number of respondents who directed comments at the competencies listed in Table 7.4. The content analysis of comments reported previously in Table 7.3 showed that consistency, player-management, and communication received numerous comments, while other competencies ranked similarly for *preparedness*, e.g., 5.3, *engages in post match activities (e.g., talk to coaches, attend post game functions)*; and, 1.3, *understands how the game is played (e.g., tactics and strategy, analyse patterns of play)*, did not warrant a single comment. Closer inspection of the *importance* and *preparedness* rankings revealed reasons for this contradiction.

Specifically, a trend emerged where numerous comments predominate when a competency is high in *importance* and low in *preparedness*. Table 7.5 details this relationship. In simplifying the interaction between *importance* and *preparedness* rankings with the number of comments each competency attracted, it can be deduced that a low *importance* ranking designates the competency as of relatively minor consequence. Therefore, competencies ranked lowly for *importance*, irrespective of the *preparedness* ranking, did not attract a

Table 7.5 Comparison of Importance and Preparedness ranking to explain comment response frequency

Competency	Importance ranking	Preparedness ranking	Comment response frequency
2.1 applies the Laws of Soccer consistently (within each game and over the season)	3	34	33
2.8 manages conflict (communication with players, use of presence and personality)	8	34	25
3.4 communicates (verbal and non-verbal) with players on and off the field	19	33	29
1.3 understands how the game is played (e.g., tactics and strategy, analyse patterns of play)	36	37	0
5.3 engages in post match activities (e.g., talk to coaches, attend post game functions)	37	36	0

significant number of comments. Alternatively, competencies perceived to be important did generate comment if not perceived to be performed to satisfactory standards.

The *preparedness* ranking of competencies also helps to explain the lack of consensual summative opinion expressed about the *overall* standard of refereeing. As discussed in the first section of this chapter, opinion was divided between positive and negative opinions. Moreover, this opinion tended to be divided on the basis of sub-group membership, with the referees, assistant referees and referee inspectors presenting favourable opinion, and the players and coaches offering negative opinion. These discrepancies can be clarified by the group differentials described in Chapter 6. As noted previously, the Rasch item estimates demonstrated a significant difference of opinion between groups for the *preparedness* perspective. In combination with the quantitative group differences for the *importance* perspective, the diversity in written opinion is likely to be a product of how competencies are perceived by each group.

In summary, findings specified various competencies which are being undertaken without adequate levels of preparation. Some of these competencies are fundamental to the referees' role. These competencies tended to attract markedly high levels of written comment, and help to explain the differences in opinions concerning the current standard of elite refereeing.

Competencies Performed with High Levels of Preparedness

Although results and discussion thus far have disclosed low *preparedness* levels for a number of competencies, the written comments and competency rankings presented evidence that a variety of refereeing competencies are undertaken with high levels of *preparedness*. These include aspects of fitness, match preparation, and one aspect of communication. The

ranking positions of each competency, according to group and overall rankings, portray a marked proficiency in the performance of these competencies (see Table 7.6).

Table 7.6 Highly ranked competencies – Preparedness

Competency	Preparedness Rankings		
	Competitive Group	Officiating Group	Overall
3.2 communicates decisions with clear hand signals	6	6	5
5.1 prepares well in advance of the match (time of arrival at ground – on time, presentation on arrival, kit prepared)	3	2	2
6.1 maintains required levels of fitness (e.g., fitness benchmarks/standards, hydration, warm up/cool down)	7	4	3

Many written comments supported this view. Specific statements include:

Physically and technically Australian referees are equal to (if not better) than in most countries. (160, referee inspector)

The fitness of referees at this level is of a very high standard and this helps all referees to control the game better. (144, assistant referee)

Majority of referees appear to be dedicated to their “profession” and display a high degree of personal preparation (fitness, dress, standards, etc). (155, referee inspector)

Superficial competencies (signals, fitness, mannerisms) are equal with anywhere in the world ... (138, assistant referee)

... in the Ericsson Cup competition, there have been some that in my opinion display a high standard of commitment and competency, including fitness. (170, referee inspector)

Emerging from these comments is an obvious, and positive, emphasis on fitness. Of those respondents who expressed a view concerning the current standard of refereeing, 22% mentioned fitness as a well-prepared competency. Further analysis revealed why this is so. Specifically, a link can be drawn between fitness, and two other competencies deemed to be high in *preparedness*. Firstly, the referee’s attention to aspects of personal health, i.e., competency 6.2, *maintains appropriate levels of personal health (e.g., correct diet, weight control)*, was given an overall *preparedness* ranking of 12. Although the relationship of diet to high levels of fitness can be blurred at times, it is less ambiguous for weight control. As the nature of elite soccer refereeing requires the referee to be aerobically fit, the consequential result of high aerobic fitness levels is improved weight control, thus presenting a fitter and more professional appearance. As such, the high *preparedness* ranking afforded maintenance of personal health was not surprising given the high standards of fitness exhibited by referees.

Secondly, the ability of referees to maintain position in close proximity to play is paramount if referees are to maximise their decision-making quality. The high *importance* ranking of competency 2.9, *moves to obtain best positions (i.e., place with best view and close enough to react effectively)*, confirms this view. Essentially, for the competency to be performed effectively (*preparedness* ranking = 11), referees rely on their fitness to ‘keep-up’ with play. This link was referred to by one player, who noted:

... being in the correct position when you make a decision gives credibility to the decision & makes players ... think twice about contradicting the decision. Being over 20m away does not. (168, player)

As can be concluded from these findings, the *preparedness* of a number of specific refereeing competencies is high. Evidence has suggested that referees are committed to their physical preparation, and this has a positive consequential effect on other related refereeing competencies. Similarly, the manner in which referees present themselves before and during a game is held in high regard.

Discussion and Implications

Overall, results presented in this section illustrate a close relationship between written comments and the ranking positions of competencies. Accordingly, Research Question 5.3: *How do qualitative comments substantiate, and elaborate upon, the quantitative findings?* can be answered through three substantial findings.

Firstly, the written comments supported the low *preparedness* ranking positions of competencies. Competencies related to consistent decision-making and inter-personal communication were implicated, and, due to the *importance* of these competencies to effective refereeing, attracted many comments. In general, the comments reflected a perceived deficiency in these two competencies. As a consequence, the findings support the conclusion that consistent decision-making and inter-personal communication are two aspects of refereeing performance requiring improvement.

Secondly, the close relationship between written comments and competency rankings were also apparent for competencies ranked highly for *preparedness*, i.e., fitness, communication (via hand signals), and match preparation. Fitness, the most important of these three competencies (ranked 9th for *importance*), attracted the most positive comment. The success of this competency can be attributed to, in part, Soccer Australia’s promotion and recruitment strategies for young career referees. Although fitness is not age specific, it was noted previously that younger referees were singled out for particular mention with respect to fitness levels. Moreover, with the increased speed of elite soccer, higher fitness demands brings the promotion of younger career referees into sharper focus, e.g.,

The game is getting faster and some of the older referees cannot keep up with play. (162, referee inspector)

Irrespective of how referees' fitness standards are maintained or developed, any subsequent reduction in this competency will lead to decreased performance levels in other refereeing competencies. Basically, the benefits of good aerobic fitness are consequential to the execution of other refereeing skills. Fitness forms the foundation for positioning, which allows the referee to make decisions close to incidents (thus maximising the information available to referees for making decisions). The push for ex-players to be recruited into elite refereeing may also be useful in regard to this aspect of referee performance. It is recognised that elite players have high levels of aerobic fitness (Reilly, 1994; Wisløff, Helgerud, & Hoff, 1998), and, as such, would not be required to 'build' this performance aspect. Rather, ex-players could maintain fitness levels similar to their playing fitness. This would be facilitated if players entered refereeing immediately at the end of their playing careers (i.e., marked reductions in fitness levels would not accrue in comparison with extended time away from regular match-play).

The third substantive link between the written comments and Likert scale responses relate to aspects of refereeing performance commented upon most frequently by the cohort. It was evident that a number of specific competencies, while being important to refereeing, are not being undertaken with adequate levels of preparation. The findings indicated that competencies implicated as high *importance/low preparedness* attracted the greatest number of comments. A number of possible reasons for this finding can be hypothesised, however, one plausible reason is that people are more inclined to discuss issues that are professionally relevant, but not performed to expectations. Nevertheless, regardless of potential reasons, the link between competency rankings and quantity of comment is distinct.

CONCLUSION

Conclusions to be drawn from the results presented in this chapter are varied, and are a reflection of the diversity of comments offered by respondents. Consequently, comments enabled a deeper insight into refereeing performance to be obtained, and provided explanations behind the findings reported in Chapter 6.

Inspection of the written comments showed an initial difference in responses with respect to quantity (from one word, to in excess of 300 words) and content. The majority of respondents gave summative evaluative opinions about the current standard of elite soccer refereeing, and expressed diverse views. A notable disagreement concerning the overall level of refereeing proficiency was evident, with opinion divided between positive and negative impressions. Further analysis showed this division was based on sub-group

perceptions, with referees, assistant referees and referee inspectors assessing standards more favourably than players and coaches.

To a certain degree, the diversity in comments explained the different perspectives and groups reported in Chapter 6. In that chapter, the perspectives of *importance* and *preparedness* were found to be statistically different, as were the perceptions the competitive and officiating groups. When these results were viewed in combination with the written comments presented in this chapter, the competencies most implicated in these differences were connected with competencies which have an immediate impact on players and their actions, e.g., consistency of rulings, appropriation of fouls, fitness, etc. It is not surprising, therefore, that players, in particular, expressed strong opinions about the performance of such competencies.

A second issue which impacted on perceptions of respondents is the criteria with which assessments on refereeing performance are made. Using the assumption that performance is based on how well important competencies are performed, a categorical determination of what constitutes ‘important’ competencies needs to be elucidated. Clear agreement between ‘assessors’ would allow judgements to be based on criteria which are, to some extent, standardised. Such a scenario was not the case prior to the present study, and was reflected through the significant differences between the competitive and officiating groups for the *importance* perspective reported in Chapter 6 (see Table 6.11), and the division of opinion expressed in this chapter.

Lastly, it became evident from comments that performance criteria can be clarified through the provision of ‘working examples.’ An obvious example emerged in this chapter. With respect to the consistency of decisions, the written comments gave detailed examples of how this skill can be judged. In particular, the issuing of free kicks and other sanctions (e.g., if one player is cautioned for tackle from behind, the next player who commits the same foul should be given the same caution), and the treatment of particular groups of players, should not vary within and across games. Referees need to be cognisant and mindful of these issues when officiating, and address them accordingly. However, notwithstanding responsibilities of the referee, the consistency examples illustrated the importance of clear and explicit understanding of performance criteria in achieving valid performance assessments.

In summary, this chapter has provided data which expanded and enriched the findings reported in Chapter 6. Reasons for the varying levels of agreement and disagreement between respondents became evident through written comments (which were both general and targeted at specific refereeing competencies). Additionally, associated refereeing issues raised in this chapter helped to explain the depth of feeling expressed – strongly at times – about the various competencies linked to refereeing performance. Implicitly, comments also

provided suggested areas of improvement that can potentially raise the overall standard of refereeing.

CHAPTER EIGHT

DISCUSSION OF RESULTS AND CONCLUSIONS

Introduction

This chapter considers the overall findings of the study in relation to the research questions and hypotheses addressed in the previous three chapters. Discussion is divided into six sections. Initially, discussion explores possible limitations imposed on the present study. Given these limitations, the second section summarises results from the investigation. With these aspects considered, the next three sections presents implications of the findings as they relate to referee training and development, referee assessment, and the utility of hybrid BARS. The final section details research initiatives generated as a consequence of this study.

POSSIBLE LIMITATIONS OF THE STUDY

The results presented in this investigation should be viewed within the context of possible limitations imposed upon the present study. These limitations can be viewed from two perspectives, namely, aspects of the research design, and characteristics associated with the research instrument.

Research Design

The integrity of the research design, and constraints imposed on the design, were discussed in Chapter 4. In particular, assessment of design validity and reliability were examined at length. Fundamental to the establishment of these design properties were two issues, namely, the scope of the study, and the nature of the sample used.

With respect to the scope of the present study, the central focus of the investigation on elite refereeing necessitated the generation of performance criteria designed specifically for the elite level of competition. To some degree, this precluded results being extrapolated widely to other refereeing contexts. However, the tight focus of the present study was consistent with underlying principles of BARS methodology. The detailed procedures of the BARS technique intentionally elicit highly specific performance criteria – a fundamental pre-requisite if criteria are going to be seen as valid representations of employee work (Stoskopf et al., 1992). Consequently, BARS data are not designed to be transferred readily to other occupations, or other levels of occupational performance (Shirom, 1988).

Given the specific focus of the research on elite soccer, a potential limitation emerged with respect to the sample. As detailed in Chapter 4, the sample included five sub-groups with vested interests in the quality of elite refereeing, i.e., referees, assistant referees, referee inspectors, players, and coaches. However, the size of each sub-group was not equivalent (see Table 4.1). To minimise the effect of differences in sub-group sample sizes, the research design incorporated specific techniques, with respect to data collection and analysis, to reduce any possible impact.

For collecting data, a range of procedures were undertaken to promote maximum collection of data from all sub-groups (particularly from those sub-groups with small sample sizes). Specifically, data from referees and coaches were collected at a pre-arranged meeting between these two sub-groups. The meeting brought together all Ericsson Cup referees, and the majority of coaches. Consequently, data were collected from 100% ($n=13$) of referees 80% ($n=18$) of coaches. Information from assistant referees and referee inspectors were collected via posting. Although this method of data collection traditionally presents less than satisfactory return rates, the motivation of these two groups to contribute to the research, plus the use of follow-up letters to encourage the return of research questionnaires, resulted in acceptable return rates (72% from the assistant referees, and 80% from the referee inspectors). Overall, questionnaire return rates from these four sub-groups provided a sound basis upon which informed conclusions could be made.

Aspects of data analysis also aided in negating the effect of differing sub-group sample sizes. In particular, two aspects stand out. Firstly, the five sub-groups were reclassified into two groups (i.e., the competitive and officiating groups) to facilitate parametric analysis. This reorganisation, based on the practical and professional alignment of each sub-group, lessened the variability in sub-group numbers that was apparent initially. Secondly, the analysis techniques used avoided potential skewing of findings that can occur when simple addition of scores is used. Specifically, the generation of items and case estimates (derived from Rasch analysis) facilitated the employment of MANOVA and ANOVA in assessing perspective and group differences. The statistical software (i.e., SPSS) used for these latter two forms of analyses takes account of differing sample sizes also, and adjusts significance levels accordingly.

Research Instrument

Limitations arising from the research instrument relate specifically to instrument piloting procedures. Piloting procedures for the present study were used to assess the design, format, instructions, and wording of the instrument (see Chapter 4). However, aspects of instrument reliability, i.e., internal consistency, were not addressed at this point. Primarily, this was because the establishment of internal consistency requires a large number of

subjects. The tight focus of the research on elite refereeing, and the resulting limitation this focus placed on potential sample size, precluded the procurement of a statistically appropriate pilot sample. This limitation notwithstanding, subsequent analysis of data verified the internal consistency of the instrument.

In summary, potential limitations were minimised through the incorporation of specific procedures. These procedures, encompassing data collection, respondent groupings, and data analysis methods, ensured the conceptual and statistical rigour of the findings.

OVERVIEW OF FINDINGS

The relational nature of the results presented in this study necessitated findings be presented in an integrated format. However, the findings can be categorised into three groups. These relate to: the nature of refereeing performance criteria identified by the hybrid BARS procedures; the suitability of BARS data for instrument development; and, the perceptions of soccer stakeholders concerning the *importance* and *preparedness* of refereeing competencies.

The Nature of Refereeing Performance Criteria

Hybrid BARS was employed in the present study to identify skills, attributes and qualities required by soccer referees to officiate elite Australian soccer. The technique, predicated on the work of Smith and Kendall (1963) and Anshel (1995), was seen to be an effective, comprehensive, and inexpensive means to collect data that described occupational performance. For the occupation of elite soccer refereeing, the process identified six performance dimensions and 38 competencies. Performance dimensions identified encompassed a range of occupational roles and responsibilities, and included: decision-making; communication; and, health maintenance. The competencies identified were also varied, and presented specific descriptions of refereeing work on and off the field of play.

With respect to the differentiation of performance criteria identified in this study, the results confirmed that officiating is a complex occupation, and requires the mastering of numerous competencies. This finding was consistent with previous studies investigating basketball umpiring (Anshel, 1995) and Touch (football) refereeing (Anshel & Webb, 1991), where a notable diversity of performance requirements were also identified.

However, the nature of each performance dimension and associated competencies identified in the present study were different to previous officiating research. The variation in findings can be explained by the focus of this investigation on an alternate sport (i.e., soccer), and, additionally, a specific level of performance (i.e., elite). Within the parameters of these two contexts, different game rules, player objectives, and game dynamics, alter the way sports are played. As a consequence, the identification of officiating requirements that were identified in this study were relevant and specific to the elite level of soccer in Australia.

The Suitability of BARS Data for Instrument Development

The performance dimensions and competencies identified from the hybrid BARS technique formed the basis of subsequent instrument development (utilised in questionnaire format). Piloting procedures reduced the number of competencies from 38 to 37 (the number of performance dimensions remained unchanged). As the questionnaire was designed specifically for the present study, no standardised validity and reliability measures were available. This necessitated the exploration of these important measurement properties.

Overall, the instrument presented acceptable validity and reliability findings. Face validity and content validity were enhanced through the generation of instrument items (i.e., performance dimensions and competencies) via the BARS process. Particularly, the use of occupational experts to identify items, and the retention of all items developed by the expert cohort, were pivotal in the verification of these two forms of validity. With respect to verifying construct validity, the 'traditional' method of establishing this measurement property (i.e., confirming the independence of performance dimensions by factor analysis) was problematic. Analysis showed the six dimension structure (identified by the expert panels) was not unique, and that other competency combinations were possible. However, Rasch analysis provided evidence that the 37 competencies were measuring one single trait related to elite soccer refereeing in Australia. Given this finding, it was concluded that one over-riding construct was being measured, and that it was appropriate for analysis to focus on competencies rather than performance dimensions. However, the value of performance dimensions should be emphasised, as they proved to be a practical way to divide this construct.

Reliability findings concerned with the instrument were impressive. Two aspects of internal consistency were determined for the present study, namely, the relationship of items to each other, and the consistency of responses across the perspectives of *importance*, *preparedness*, and *improvement priority*. The former was measured using Cronbach's alpha coefficient, and was reported at 0.96. The latter was assessed using Pearson's Correlation Coefficient (r). If respondents were responding to each perspective as intended, the difference between the *importance* and *preparedness* responses would have been reflected in *improvement priority* scores. Calculation of Pearson's r at 0.93 (significant at $p < .001$) provided evidence to confirm the intended response pattern.

Perceptions of Competency Importance and Preparedness

Analysis of Likert scale responses for the *importance* and *preparedness* perspectives was undertaken using frequency analysis and Rasch scaling. The latter form of analysis provided the clearest indication of how each competency was perceived. Specifically, item estimates derived from the *importance* data provided substantive evidence that a stable hierarchy of

competency importance exists. Most notably, competencies associated with decision-making and rule application were viewed as most important.

Similarly, a quantifiable continuum was evident with respect to the *preparedness* with which refereeing competencies are undertaken. However, it was notable that various competencies ranked highly for *preparedness* were not ranked highly in terms of *importance*, e.g., *effectively uses the whistle (e.g., volume, tone, timing, length, player reaction)*. As a balance to this finding, a number of *important* criteria were perceived to be undertaken with low levels of *preparedness*, e.g., *applies the Laws of Soccer consistently (within each game and over the season)*. Examples such as these contributed to a significant difference ($p < .001$) between perspectives.

Moreover, perceptions between the competitive group (players and coaches) and the officiating group (referees, assistant referees, and referee inspectors) within each perspective were found to be significantly different ($p < .004$ for *importance*; $p < .001$ for *preparedness*). Reasons for these differences can be found, to some extent, by rankings given to criteria within each perspective. For *importance*, players and coaches tended to rank performance criteria associated with the interpretation of rules and game incidents highly. Conversely, referees, assistant referees and referee inspectors viewed criteria that required less interpretation, e.g., refereeing mechanics, as most important. Similar variations in group perceptions were apparent also for the *preparedness* data.

Written comments supplied by respondents supported the Likert scale responses. Overall, a clear differentiation in opinion emerged between the officiating and competitive groups when asked to comment on the current standard of elite refereeing. The officiating group viewed refereeing performance in a positive light, while the competitive group were more critical in their perceptions. In particular, this group provided consensual opinion that decisions needed to be more consistent, and communication between referees and players required considerable improvement. However, overall, comments presented a positive impression of the fitness levels of referees (especially the younger referees), and the performance potential of the current cohort of younger referees.

Aside from the performance aspects of refereeing, possible mechanisms to improve overall refereeing standards emerged from the data. These included the recruitment of ex-players into refereeing, and improvements in the assessment and evaluation process of referees. The provision of such data suggested the respondents also saw refereeing performance in broader terms. In doing so, they demonstrated inter-relationships between performance, performance assessment and training. By implication, this deeper level of analysis by respondents provides direction for improving refereeing (and the overall spectacle of the game).

Summary

The findings of the present study provided a unique perspective on the officiating occupation. For the first time, a comprehensive analysis of refereeing requirements, as they relate to elite soccer refereeing, have been presented. Additionally, the study demonstrated that some competencies are of greater importance to the refereeing role, and that these need to be undertaken with higher levels of preparation.

IMPLICATIONS FOR THE TRAINING AND DEVELOPMENT OF ELITE SOCCER REFEREES.

Explicit performance criteria provide a sound basis for training curricula (Rutherford, 1995). In deference to Rutherford's assertion, two substantial implications for the training and development of elite soccer referees emerged from the present study. These are the diversity of referees' work, and the relative emphasis in training programs on varying aspects of referee performance.

The Diversity of Refereeing Requirements

This study confirmed the variety of tasks referees are required to undertake. As detailed in Chapter 3, numerous on-field roles and responsibilities (e.g., decision making), plus expectations and obligations carried out away from the competitive arena (e.g., maintenance of fitness) were identified as being essential for effective soccer refereeing. Such diversity was not surprising, particularly given the wide range of performance requirements noted in the relevant literature (see, for example, Anshel, 1995; Anshel & Webb, 1991; Clegg & Thompson, 1993; National Officiating Program, 1997; Weinberg & Richardson, 1990). Given this multitude of requirements, training and development programs need to encompass, and articulate with, all essential skills and attributes. Training programs that are relevant to the needs of the sport, and the official, should be seen as a fundamental first step in raising refereeing standards.

Further, the recognition of all requisite refereeing performance criteria needs to be undertaken using a systematic, valid and reliable method. However, it is apparent from the current officiating literature that this is not the present *modus operandi*. Only basketball (Anshel, 1995) and Touch football (Anshel & Webb, 1991) have made serious attempts to clarify performance criteria. Notwithstanding this systemic problem, the present study has demonstrated that hybrid BARS, through its extensive occupational analysis process, provides a viable and manageable technique for identifying essential refereeing performance criteria. Moreover, as noted by Anshel and Webb (1991), performance criteria generated by BARS technique provide a framework for prospective or inexperienced referees to model their professional practice.

The Focus of Training Programs

Although all performance criteria are essential to some degree, it is impractical to assume all criteria require the same prominence within training programs. As such, training curricula should take into account the relative emphasis to be placed on each performance criteria. Implicitly, the most important criteria are normally those performed frequently, or are of such fundamental consequence to the successful execution of occupation performance that their importance is broadly recognised (Bradley & Pursley, 1987). The implication, therefore, is that the greatest training emphasis be given to the most important criteria. Failure to observe this concept may result in finite training resources being allocated to criteria that have relatively diminished importance to the occupational role.

Notwithstanding the significance of this implication, empirical examination of the relative importance of refereeing performance criteria in soccer has not been undertaken prior to this study. However, through Rasch scaling, the present study established a clear *importance* continuum for elite soccer refereeing performance criteria. Similarly, the present study also determined which criteria are being performed with inadequate levels of preparation. In combination, these analyses bring training and development needs into sharper focus.

More specifically, from this investigation, two areas of refereeing performance emerged that warrant greater prominence in training and development programs, i.e., the consistent application of laws during a game, and the nature of a referee's communication with players. Consistency is probably the most problematic of these two issues to address, mainly due to the diversity of contexts in which referees work. Although differing game contexts do not foster repeatability in rule application and decision-making (Wilson, 1986), strategies are available to referees for improving this aspect of their work (Murphy et al., 1995). Examples of these strategies include:

- detecting when attention levels decrease so 'triggers' can be developed to regain necessary concentration;
- maintaining appropriate levels of fitness (fatigue can impair the decision making process);
- preparing for the environment in which the event will occur. This can range from the influence of parochial crowds to the speed at which the game is played (usually, the higher the standard, the less time is available to the referee to make a decision); and,
- retrieving pertinent information from 'working' and 'long-term' memory. Obviously, if decisions are to be consistent, the referee must be cognisant with decisions that have accompanied the same or similar game situations.

In light of these suggestions, officiating bodies should be mindful of such strategies, and take steps to implement them into their respective officiating curriculums. While it could be argued that attention to the strategies may not remove *all* inconsistent rulings, they have the potential to minimise the occurrence of ambiguous decisions. As stated by Murphy et al. (1995, p. 21):

Accurate and consistent decision making, though desirable and from all officials at all times is in reality very difficult to achieve. However, with regular practice, errors and variability in performance can be reduced.

With respect to communication issues, the way a referee imparts decisions (referred to as paralanguage, see Evans, 1994) impacts on how players react to other players and to the referee. This relationship was observed by Anshel (1989, p. 32), who noted, “the manner in which an official communicates with other participants during a contest is important to maintain the proper competitive environment.” However, findings from the present study suggest that communication of this type is virtually non-existent, and, when it does occur, is of poor quality. In addition, evidence indicated that a more thorough approach to this performance criteria is required in referee training programs if standards are to be improved.

From a more general perspective, training and development programs require flexibility to accommodate the varying abilities of prospective referees. This is particularly important if referees come to the profession through differing recruitment pathways. Given this recognition, all referees do not need to follow rigid curriculum content and training procedures. For example, training curricula should vary according to the ‘entry level’ status of referees (i.e., younger career referees follow a different program to that of ex-players). The advantages of diversifying training programs should be apparent through improved retention of referees (due to the relevance of the curriculum), and improvement of overall refereeing standards (due to the improved standard of refereeing within the pool of elite referees).

In a related point concerning ex-players, the recruitment of ex-Ericsson Cup players into elite refereeing has had marginal success. No Ericsson Cup refereeing panel has, or has had, any former Ericsson Cup players. This circumstance is probably to the detriment of overall refereeing standards, as recently retired elite players possess higher than average speed, endurance, and knowledge of play (Rontoyannis, Stalikas, Sarros, & Vlastaris, 1998). These factors, according to Rontoyannis et al. (1998), qualify former players “as the best candidate to become a successful football [soccer] referee” (p. 213). However, various issues, including relatively poor remuneration, are cited as reasons why elite players seldom enter the refereeing profession (Rontoyannis et al, 1998).

The issue of remuneration may be addressed to some degree by a move to the full-time employment of referees. If such a move were to occur, the long-term employment prospects

of referees require consideration. Given the current retirement policy for elite referees, these people would be seeking a new career path at 45 years of age. Similarly, if referees could continue in the Ericsson Cup indefinitely, it is probable that reduced fitness and form would lead to exclusion from the national referees panel before the 'normal' societal retirement age, i.e., 60-65 years of age. Again, this scenario would leave referees seeking gainful employment in other occupations at a latter stage of life. These concerns, while hypothetical in the current soccer environment, are pivotal to successful full-time officiating, and the financial security of referees.

Nevertheless, there are three important implications associated with salary issues for referee training and development. Firstly, a valuable pool of potential referees, i.e., ex-players, are not entering the refereeing profession. As a consequence, the soccer understandings that this group bring to refereeing is lost. This not only weakens the talent pool by their direct absence, but precludes the benefits of cross-education with career referees. Secondly, for referees to be full-time, people who are responsible for the professional development of referees need to be employed. As pointed out by the Director of Referees for Soccer Australia (pers. comm. Gary Power, 7 July 1999), there is little use in having full-time referees if there is nobody to coach them on a full-time basis. The third implication is a function of the second. For referees and their coaches to be employed full-time, a monetary cost is involved. Without the necessary fiscal resources for wages and training support, a move to full-time refereeing is problematic.

Summary

The performance criteria identified in the present study provides referee educators with a framework for developing a more focused curricula. The fundamental benefit of the framework is that it details the most important aspects of refereeing performance, thus facilitating training programs that are predicated on relevant 'industry' needs.

IMPLICATIONS FOR THE ASSESSMENT OF REFEREES

Masters and McCurry (1990) noted that judgements about any performance should be made against valid and reliable criteria (it was these two criteria that first led Smith and Kendall (1963) to develop BARS). In combination with such criteria, people who make judgements need to be cognisant with the criteria's meaning and intent. Moreover, assessors are obliged to understand how criteria apply to the occupation under review. A recognition of the symbiotic relationship between performance criteria and assessors is essential if assessment procedures are to be seen as legitimate appraisals of occupational performance. This is particularly pertinent if assessments are to be used as a basis for promotion, demotion, and salary rises (Murphy & Cleveland, 1995). It is within this dual relational context, between performance criteria and assessors, that implications for referee assessment are examined.

The Appropriateness of Assessment Criteria and Assessment Reporting Formats

Currently, the assessment of elite refereeing performance in Australia is predicated on criteria developed by FIFA, the governing body of world soccer (see Appendix 27). These criteria are performance-domain oriented, with minimal detail provided to clarify the meaning and intent of each domain. At face value, a concise format is apparent. However, the format does little to foster unambiguous interpretation of performance. For example, one FIFA criterion evaluates the impact of referee 'personality,' with referee inspectors asked to determine if the referee's personality "assisted in their [the referee's] control of the game." Notwithstanding the inherent concerns in assessing personality, two obvious assessment problems emerge from this example. Firstly, the referee inspector makes evaluative judgments at a considerable distance from the field of play (normally in the grand-stand). Such minimal proximity to the game renders the reliability of any judgement questionable. Secondly, the determination of a 'suitable' personality is vexed. During the course of a game, referees are required to interact with numerous personalities (at least 22 players and two coaches), with various game situations promoting different aspects of individual personality. The expectation of a referee to align his personality with all these personality variables is somewhat unrealistic. As noted by Conway and Ellison (1995, p. 104), "measuring traits, personality and global satisfaction had proved to be less related to one's performance than to other issues."

To compound this problem further, three additional concerns are apparent. Firstly, performance criteria are inconsistent across various assessment reports. Currently, Soccer Australia triangulates performance data by having assessment reports completed by referee inspectors (Appendix 27), club representatives (Appendix 28), and referees (see Appendix 29). However, all three report formats do not present criteria consistent with respect to the following: the number of criteria; the type of criteria; and, weightings applied to each criteria. Accordingly, meaningful triangulation among assessors is not possible, thus limiting the breadth of conclusions that can be made across assessment formats.

The remaining two concerns relate specifically to determining a summative mark or grade that reflects referee performance. The first of these concerns centres on the 'score,' allocated by the assessor, for each performance criterion. For referee inspector reports, criterion are allocated a numerical score on a 1-to-10 continuum, with each score having a related description, e.g., a score of 5 = 'Fair.' (As a side issue, inspectors are required to differentiate performance descriptions that are closely aligned, e.g., 'Outstanding' and 'Excellent' represent scores of 10 and 9, respectively. Such attempts at differentiation can be problematic due to possible varying interpretations of the words 'Outstanding' and 'Excellent.')

Other assessment formats present different scoring options. For example, club

reports use four classifications, while referee self-appraisal reports use five scoring classifications. Given the need for judgements to be based on clearly defined and consistent performance levels (Masters & McCurry, 1990), reliable performance assessment using the current format is unlikely.

The present study has provided a means to overcome these concerns. Specifically, the number of criteria that can be used to assess performance is increased to 37. This increase allows criteria to be written in more specific terms, thus avoiding many of the ambiguities common to performance assessment. As a consequence, criteria tend to be less ephemeral, more accurate and defensible in assessment contexts. The advantages of increased criteria numbers is noted by Oppenheim (1992, p. 147), who stated,

Sets of questions [i.e., items] are more reliable than single option items; they give more consistent results, mainly because the vagaries of question wording will probably apply only to particular items, and thus any bias may cancel out.

Nevertheless, not all the criteria identified in the present study relate to on-field performance. Specifically, seven criteria can be considered to be performed away from the competitive context of the game. As such, it needs to be determined how extensive performance evaluations are required to be. When doing so, referee administrators should be mindful that performance criteria developed in the present study used a diverse range of experts, and were subsequently validated by a similarly diverse, but more extensive, cohort of soccer stakeholders. The implication is that administrators may need to consider incorporating such criteria into performance evaluation given the wide support of these criteria from the soccer community.

The final concern with the current assessment system relates to the 'weighting' of scores. For example, within the referee inspectors' report, the 'Fitness and Positioning' criterion is given a weighting of two, while the 'Interpretation and Application of Laws' criterion has a weighting of three. When compiling the referee's summative grade, scores for each criterion are multiplied by the appropriate weighting and then summed. This process implies that some criteria are more important than others when determining the overall quality of referee performance. However, there is no empirical basis for the relative magnitude of each weighting (National Director of Referees, pers. comm. 7 July 1999). Moreover, the weighting system is not consistent across all assessment formats (i.e., only inspectors' reports incorporate the weighting system). Discrepancies such as these magnify concerns with current rating formats, and, once again, minimise the potential for assessment data to be triangulated.

In contrast, the present study has provided an empirical basis for the weighting of performance criteria (the weightings are supplied through item estimates derived from Rasch analysis). This allows for the most important criteria to be given the highest weighting, and

the weighting of other criteria adjusted accordingly. Moreover, the weightings provided in the present study are a result of various soccer stakeholders having input into the 'weighting' process. The input has the potential for greater ownership of both criteria and weighting across the soccer community, and, as such, represents a unique opportunity to unify assessment formats for all assessors.

The Referee Inspection Process

Notwithstanding the importance of clearly defined, unambiguous and consistent performance criteria, the expertise of assessors is similarly pivotal to legitimate performance appraisal (Rutherford, 1995). This notion was given prominence in the written responses discussed in Chapter 7. The point was made that some referee inspectors lack the understandings required to assess refereeing performance reliably. With this concern in mind, a number of implications arise for the selection of inspectors. These are based on suggestions offered by Preston and Kennedy (1995), and include:

- assessors being expert in the field they are assessing;
- assessors undertaking training in the techniques of observational assessment;
- assessors gaining extensive experience in assessment, so they become expert in this function; and,
- assessors coming together to share their standards, discuss cases, and converge on some consensus about the standards of each performance criteria.

Given the benefits that can be derived from effective inspection and assessment procedures, the importance of reforms to the inspection process should not be ignored. Fundamentally, the responsibility for improving inspection standards lies with Soccer Australia. The organisation needs to take steps to ensure inspection procedures are thorough and standardised.

Concerns relating to the refereeing experience of inspectors are more difficult to overcome, particularly in the short term. Comments indicated that refereeing experience at, or close to, the Ericsson Cup level should be a pre-requisite for the role of referee inspector. However, only a limited pool of ex-Ericsson Cup referees are available, and of this cohort, other commitments, such as family and occupational responsibilities, have the potential to preclude them from inspector duties. Nevertheless, any moves to recruit former Ericsson Cup referees into referee inspecting presents advantages to the assessment process. As noted by Rutherford (1995, p. 105),

Unless an assessor has had a background identical ... to that of the candidate then there is every possibility that a wide range of evidence will be put before

him or her of which some, if not a great deal, he or she may not at first recognise or understand.

Summary

The performance criteria and weightings developed in this investigation have, for the first time, provided a basis for referee assessment to be grounded in sound methodological and statistical foundations. This study has shown that the criteria have established reliability and validity measures, and, as such, present an alternative assessment structure to that currently in use. However, it is recognised that the criteria, along with respective weightings, have not been applied in practical assessment settings. Implementation of criteria in this manner is fundamental to establish the utility of criteria for assessment purposes (this issue is addressed later in this chapter, see *Recommendations for Future Research*).

IMPLICATIONS FOR BARS

Findings from the present study raised two important implications for the development and use of hybrid BARS. Specifically, implications focus on the 'external validation' process, and the process used to classify competencies into performance dimensions.

The 'External Validation' Process

The present study used 'external validation' to assess the degree of support the wider soccer community has for the occupational competencies identified by the BARS expert panels. This was different to previous hybrid BARS studies which determined support for performance dimensions only (Anshel et al., 1987; Anshel & Webb, 1991; Anshel & Weinberg, 1995; Jessup, 1994; Moore & Webb, 1995). Moreover, the suitability of performance dimensions in these previous studies were assessed using a technique that lacks statistical rigour, i.e., cumulative frequency analysis. Accordingly, this study sort to determine the suitability of hybrid BARS for identifying essential performance criteria through more sophisticated statistical techniques, i.e., Rasch analysis. As a side issue, this provided a unique comparison between the merits of Rasch scaling and traditional methods of analysis.

Rasch analysis provided evidence that only one criteria, i.e., *engages in post match activities (e.g., talk to coaches, attend post game functions)*, was not seen to be important for effective refereeing performance. This finding is in contrast to the four items labelled 'unimportant' using the arbitrary cut-off criteria utilised in cumulative frequency analysis techniques. The discrepancy cautions researchers against accepting the findings of frequency analysis procedures as a *fait accompli*. As an alternative form of analysis, Rasch scaling proved to be more suitable given the provision of item estimates, standard deviations, and the use of these parameters for determining relationships among items on a continuum.

A second issue related to the 'external validation' process was the use, in data collection, of groups with different soccer roles. The use of an assorted cohort in the present study represented diversification in the external validation process not apparent in previous studies (see, for example, Anshel, 1995; Anshel et al., 1987; Anshel & Webb, 1991). Given the restricted samples used in these studies, it can be argued the validity of the external validation findings are somewhat compromised. However, in this investigation, diversity of opinion was seen to be fundamental to establishing the credibility of the findings, and consistent with the diversity of personnel used for performance dimension and competency identification.

As a consequence of obtaining data from a range of soccer stakeholders, findings from this investigation showed that the different sub-groups were not always in agreement. This finding demonstrates that data from homogenous samples (i.e., referees only) are not necessarily representative of all occupational stakeholders. The implication for future hybrid BARS studies is that researchers should be mindful of varying opinions, and consider collecting data from a range of relevant stakeholder groups.

The Integrity of Performance Dimensions

The present study sought to obtain empirical verification of performance dimensions identified during hybrid BARS development. To this point, the practice of classifying competencies into performance dimensions was accepted to be valid. This acceptance prevailed despite evidence from traditional BARS procedures that this was not always the case (Bernardin, 1977; Dickinson & Trice, 1977; Kafry et al., 1979; Keaveny & McGann, 1975). Accordingly, this study undertook to verify the classification process through factor analysis and Rasch scaling.

In the case of factor analysis, the technique was able to provide an initial indication that the performance dimension structure was not supported using statistical analyses. This was the case when the technique was applied to all three perspectives from which data were obtained. Similarly, Rasch analysis provided evidence that single traits, within each perspective, were being measured. Findings such as these indicate that alternative competency groupings are possible, and the dimension structure derived from the BARS process is not necessarily unique.

A probable reason for these results can be found in the sequence of steps followed in the dimension development process. As described in Chapter 3, the present study applied the hybrid BARS technique, and identified competencies prior to dimension development. Specifically, dimensions were 'named' as a result of like competencies being grouped together into seemingly homogenous categories. This sequence had its foundations in traditional BARS literature (e.g., Bradley & Pursley, 1987; Campbell et al., 1973), where it was seen to be beneficial to keep experts focused on specific actions rather than on global

traits. However, this sequence also represented an alteration from the sequence proposed originally by Smith and Kendall (1963) and Landy and Guion (1970). In these two studies, broad performance dimensions were identified first, then descriptors identified which were specifically representative of each performance dimension. Therefore, the implication for hybrid BARS is to contrast the effect of generating performance dimensions prior to competency development, i.e., the sequence used by Smith and Kendall (1963) initially, with the process used in this study. Such an analysis may provide deeper insights into the relationship between performance dimensions and competencies (see *Recommendations for Further Research* for more details).

Summary

Given the implications presented in this section, development procedures used in the hybrid BARS technique need to be reassessed. As the technique is still in its relative infancy, procedural and analytical implications will continue to arise as they did for traditional BARS in the 1970s. However, any implications that do arise should not be interpreted as a weakness in the hybrid BARS process. Rather, the implications should be viewed as a necessary pre-requisite for establishing further the process' methodological credentials.

RECOMMENDATIONS FOR FURTHER RESEARCH

The findings presented in this study have raised additional issues worthy of further investigation. Five stand out, and are listed below.

1. The statistical support of performance dimensions, when competencies are generated to represent specific performance dimensions, needs to be assessed. As noted in the previous section, the procedures undertaken to generate performance dimensions in the present study necessitated competencies be identified initially, then classified into performance domains. This approach, consistent with hybrid methodology (Anshel et al., 1987; Anshel & Webb, 1991; Jessup, 1994), was unable to produce performance dimensions that were supported by statistical analyses.

To produce dimensions that are supported statistically, two possible procedural routines are worth exploring. Firstly, performance dimensions should be developed initially, followed by the generation of competencies that are representative of specific dimensions. This sequence generates competencies aligned to each performance dimension, and may improve the likelihood of sustainable factor structures being generated.

A second alternative is to have an instrument comprised of competencies only. The identification of performance dimensions occurs after data collection, i.e., factor analysis and Rasch scaling can be utilised to determine the existence of underlying

traits or constructs. A technique such as this is similar to that undertaken previously by Stoskopf et al. (1992) using traditional BARS techniques. In their study, a stable three factor solution was derived from an initial 27 items related to nursing performance using factor analysis. Although it could be hypothesised that results such as those reported by Stoskopf et al. (1992) were a 'one-off,' additional research will clarify the success of using the data (supplied by survey respondents), rather than expert panels, for generating performance dimensions.

2. The utility of competencies for evaluating *individual* performance needs to be determined. Although refereeing performance was assessed to some degree in the present study (by comparing the *importance* and *preparedness* ranking of competencies), the assessment was predominantly holistic. The performance of individual referees, in specific performance contexts, was not addressed.

For individualised assessment to be valid, a first step is for performance criteria to be assessed for inter-rater reliability. The establishment of this measurement parameter is fundamental to the integrity of assessment instruments. The one hybrid BARS study that has assessed inter-rater reliability (Anshel, 1995) indicated acceptable levels of agreement between raters (0.88). However, the study of Anshel assessed agreement using performance dimensions, rather than specific competencies.

For determining levels of inter-rater reliability (utilising the competencies generated in the present study), it is recommended that inter-rater reliability be computed using pairs of assessors from the same soccer background, e.g., pairs of referee inspectors, or pairs of coaches. This design is appropriate because reliability findings using varied assessment groups may be difficult to interpret, particularly given likely systemic differentiation in the intensity of the 'critical eye' applied to performance. This problem notwithstanding, if assessor pairs comprised of individuals from differing sub-groups are used, some indication of agreement between all assessment pairings may be achieved through a comparison of relative scores given to each criteria. Criteria can be ranked for each pair, then correlations made with other pairs using a statistical technique, such as Pearson's Rank Correlation. Strong correlations would support further the inter-rater reliability of performance criteria.

Aligned with this process should be investigations surrounding the integrity of hybrid BARS based instruments with respect to leniency error and halo error. Controlling these two forms of measurement problems consumed numerous research projects using traditional BARS techniques (Bradley & Pursley, 1987; Landy, 1985; Leap & Crino, 1993; Piotrowski, Barnes-Farrell, & Esrig, 1989; Schwab et al., 1975). Inconclusive findings resulted from these studies. However, the hybrid version of BARS has not been subjected to similar scrutiny. Detailed studies need to ascertain the

effect that leniency and halo have on the newer assessment technique prior to hybrid BARS being accepted as a legitimate assessment tool.

3. The effectiveness of criteria developed in the present study for providing feedback needs to be determined. Literature pertaining to traditional BARS has highlighted improvements in feedback quality when BARS-generated data are used for assessment purposes (Hom et al., 1982; Shirom, 1988). This was due to the assessment criteria being grounded in examples of occupational performance that are real and reflect experienced work situations (Shirom, 1988).

However, the present study identified competencies that were based on more than just observed behaviours. Competencies were also related to occupational knowledge, understandings, and attitudes. Given the altered nature of the competencies, additional research into hybrid BARS data, for feedback purposes, is necessary. Specifically, it is recommended that research focus on the impact of feedback on referee training and development. Collection of qualitative data from those most implicated in the feedback–training loop, i.e., referees, referee inspectors, and referee coaches, should provide valuable insights into the usefulness of BARS data for feedback.

4. The possibility of generic performance criteria existing across officiating standards needs to be assessed. Although this study outlined the necessity, and advantages, of identifying performance criteria defined specifically for one occupational group, it may be the case that many criteria are common to soccer refereeing, irrespective of the competitive standard of play. Previous work of Anshel (1995) and Anshel and Webb (1991) alluded to this possibility.

Intuitively, the existence of generic competencies within one sport would appear to be common-sense. However, no research with this focus has been conducted on the necessary performance criteria at junior, amateur, local association, or state-league levels. Research examining the possibility of generic competencies has the potential to provide valuable information for developing a global perspective on the training and education needs of soccer referees.

CONCLUSIONS

The present study examined the occupational requirements of elite soccer referees. This was carried out using a model based on identifying relevant behavioural work descriptions (i.e., BARS). However, the study expanded competency identification to also include knowledge, attitudes, and understandings related to occupational performance. This expanded view of competency development represented a more contemporary approach to defining

occupational requirements. In doing so, the study attempted to avoid problems many researchers have experienced using behavioural models.

In particular, the findings of the study allowed for the work of elite soccer referees to be clarified. It confirmed the diversity of occupational requirements that were formerly suspected, although never proven through research. The diversity of performance requirements identified were consistent with studies in other sports (Anshel, 1995; Anshel & Webb, 1991), and with generic officiating competencies promoted by the National Officiating Program (1997). However, the competencies identified in the present study also demonstrated the unique requirements of soccer refereeing at the elite level, and confirmed the assumption that officiating performance requirements need to be developed specifically for each individual sport.

One finding the study has highlighted is the need for a variety of occupational stakeholders to be part of the competency development process. Given that the execution of occupational performance impacts on closely-aligned occupational sub-groups, the opinion of these affected sub-groups require consideration. The sub-groups have the capacity to provide perspectives on the occupation which may go unnoticed if competency development remained the sole responsibility of people within the profession, e.g. referees only commenting on refereeing. The significant differences between different groups of stakeholders, as demonstrated in this study, is testament to this possibility.

Lastly, the study gave prominence to problems that can occur when using multiple forms of assessment, and, moreover, using assessment criteria that lack validity and relevance. The need for standardised criteria, that reflect accurately the nature of refereeing work, is a fundamental starting point for assessment. Anything less compromises the assessment process, and hinders potential improvement.

POSTSCRIPT

As a consequence of this research into elite soccer refereeing, and following the release of findings from this study at national and international conferences, a research project similar to the present study has been commissioned by the Australian Rugby Union. The project takes into account the contextual framework and procedures established in this study, but differs in that it examines refereeing performance at differing tiers of rugby. In doing so, it is anticipated the project will provide a training and education framework for the development of rugby referees. The project is due for completion in mid-2000.