

## CHAPTER THREE

# PRELIMINARY STUDY

### Introduction

The previous two chapters provided an overview of issues which are worthy of further investigation. In particular, Chapter 1 addressed problems associated with defining and clarifying tangible occupational criteria required by sports officials. Further, it examined the link between the performance of these criteria and challenges currently confronting the officiating profession. In Chapter 2, a procedure, which has been successful in identifying performance criteria across a variety of occupations, was examined. Discussion articulated how criteria can be defined specifically, yet be diverse enough to encompass an entire occupational domain.

From these chapters, a variety of research themes emerged. This chapter addresses one of these, namely, the identification of refereeing performance criteria essential for elite soccer. In general terms, the chapter details how hybrid BARS techniques were adapted and utilised for performance criteria identification of elite soccer referees in Australia.

Specifically, the chapter is divided into four sections. The first section provides background for the preliminary study, including a definition of pertinent key terms and the planning considerations that preceded the identification of criteria. The second section describes specific procedures undertaken in identifying the performance criteria. In particular, the role of the expert panels is the major focus of the discussion. The third section reviews the criteria (results) derived from section two, and investigates each dimension (and related competencies) in detail. The final section discusses the implication of the results, plus compares the results with other studies investigating performance requirements of officials in other sports. In doing so, the final section addresses the following research questions:

- 1.1 What performance dimensions and competencies, identified using hybrid BARS development procedures, are essential for elite soccer referees?
- 1.2 Are performance dimensions and competencies reflective of officiating performance criteria in other sports?

## BACKGROUND TO PRELIMINARY STUDY

As noted in Chapter Two, the work of Anshel (1995) and Anshel and Webb (1991) represented the only known attempts to delineate empirically the key performance criteria of officiating. Mirroring this circumstance, literature related to specific performance criteria of *elite* soccer referees is similarly unsubstantial. To address this short-fall in officiating research, the present study employed an adapted hybrid BARS methodology, as described in Chapter Two, to determine the performance requirements of elite soccer referees. Two overriding reasons are advanced for utilising this technique. Firstly, it has been established that BARS methodology – whether hybrid or traditional versions – encompasses extensive occupational analysis (Anshel, 1995; Borman & Dunnette, 1975; Grussing et al., 1979; Jacobs et al., 1980; Schwab et al., 1975). Such analysis is of fundamental importance to the present study, particularly given the study's focus on occupational performance criteria. Secondly, these criteria facilitate exploration of relevant research themes through additional qualitative and quantitative analytic techniques.

The hybrid BARS procedures used in the present study were based on those described first by Anshel et al. (1987), and adopted subsequently, with minor modification, in a number of other studies (Anshel, 1995; Anshel & Webb, 1991; Jessup, 1994; Webb et al., 1994). However, before proceeding with explanation of how the technique was applied in the preliminary investigation, the context of the preliminary investigation is examined in this section. Discussion is devoted to defining terms which framed the preliminary study, and to providing an overview of the planning considerations which were fundamental to the hybrid BARS process.

### Definition of Terms

A number of terms specific to the preliminary investigation require clarification, namely: *competency*; *performance dimension*; *performance criteria*; *elite*; and, *Ericsson Cup*. Definitions of each of these terms is provided below:

1. *Competency* – a definable knowledge, understanding, skill, or attitude made explicit through actions (Bartlett, 1992). Competencies may be simple or complex. This variation in complexity is illustrated by Bartlett (1992) in the context of classroom teaching. Simple competencies included tasks such as writing on the white-board, while competencies such as lesson-pacing were seen to be complex.
2. *Performance dimension* – a term or statement that represents a cluster, or range, of competencies. For the purposes of this investigation, performance dimensions reflect “the necessary, desirable skills and tasks for meeting the criteria for job effectiveness” (Anshel, 1995, p. 6).

3. *Performance criteria* – a collective term used to encompass performance dimensions and competencies.
4. *Elite* – participation, or involvement, in soccer at the Ericsson Cup standard.
5. *Ericsson Cup* – the premier domestic soccer competition in Australia. It is played on a home-and-away basis between competing clubs. It is conducted over an eight-month period, beginning in October and finishing in May. In the 1997-98 season, the Cup featured 14 clubs located in Brisbane, Newcastle, Sydney ( $n=3$ ), Wollongong, Canberra, Gippsland, Melbourne ( $n=3$ ), Adelaide ( $n=2$ ) and Perth (Soccer Australia, 1997a).

These definitions, especially those for *performance dimension*, *performance criteria*, and *elite*, are not necessarily dictionary definitions. Similarly, the definition of *competency* is not definitive, as definitions can vary within the competency-based literature (for example, see Anshel, 1995; Bartlett, 1992; Leap & Crino, 1993; Preston & Kennedy, 1995; Rutherford, 1995). However, all definitions used in the present study can be considered ‘working definitions’ within the context of this investigation. As such, they provide a framework in which the hybrid BARS process can proceed.

### **Planning**

The remainder of this section is devoted to delineating fundamental organisational tasks that preceded the BARS process. Central to this process was a one-day (six-hour time frame) meeting of soccer experts, who were charged with the responsibility of identifying relevant performance criteria. For the meeting to proceed smoothly and to maximise productivity, a number of pre-meeting considerations were addressed. In particular, these included the specification of the target population, selection of experts for panel membership (key soccer stakeholders), and the identification of suitable support personnel.

### **Identification of the Target Population**

The importance of a well-defined target population is paramount, as the occupational performance criteria identified through the BARS process are specific to an occupation and the organisational level of the occupation. Consequently, the need for a tight focus on occupation, and occupational level, meant the target group needed to be defined carefully (implicitly, this consideration also improves the face validity of resulting data, see Anshel et al., 1987). Given these parameters, the target group for the present study was identified as elite soccer referees, i.e., those people who are given full authority to administer the laws of soccer during Ericsson Cup soccer matches.

### **Selection of Experts for Panel Membership**

The primary role of the 'expert panel' is to develop performance criteria that reflect the role under review (Cyrus, 1979). For the purposes of this study, 'experts' were defined as individuals "with extensive knowledge, familiarity, expertise, or previous experience with the criterion job or task" (Anshel, 1995, p. 6), and representative of "various roles, knowledge, past experience and needs of the target profession" (Anshel, 1995, p. 14). Accordingly, experts for the present study were required to be: experienced in their respective roles at the highest level of Australian soccer; and/or representative of such people; and, cognisant of current officiating issues.

Additional consideration was given to diversity in expertise, i.e., the inclusion of soccer stakeholders aside from referees. Diversity of this nature is not without precedent. For example, when identifying the competencies for Touch officials, Anshel and Webb (1991) used experts from a variety of Touch stakeholders, and included: state and national referee's directors; referees; players (state and 'park' level); technical and development officers; and, an academic from a related field of study. Similarly, in an investigation of classroom discipline and management, Jessup (1994) used educational consultants, principals, special education teachers, and regular classroom teachers, to comprise two expert panels. The theoretical assumption underpinning such diversity was to ensure the pertinent occupation is examined in detail from many perspectives, thus improving the content validity of the resulting data.

Given these precedents, it was appropriate for the present study to draw experts from a variety of soccer stakeholders. Experts included: Ericsson Cup referees; Ericsson Cup assistant referees; Ericsson Cup referee inspectors; Ericsson Cup players; National Directors of Coaching (current or former); representatives of the Australian Coaching Council; and, academics from a cognate area. In deference to recommendations from BARS literature (Anshel, 1995; Moore et al., 1997), two representatives of each stakeholder group was required. As such, two panels were designated, with expertise consistent across each panel, e.g., one player per panel, one referee per panel, etc.

Three comments need to be made relating to the composition of each panel. Firstly, it was desirable to have current Ericsson Cup coaches on each panel, however, such a proposition was not feasible. This was due mainly to the availability of current coaches to attend. Specifically, there were only three current Ericsson Cup head coaches in the city where the meeting was to occur (Sydney). Although all coaches expressed an initial willingness to participate, it was logistically impossible to find mutually agreeable dates for any two coaches to be available.

In an attempt to overcome this constraint, former Ericsson Cup coaches were also considered, but once again, their lack of availability, coupled with the need to standardise recent experience across the panel, also made this option inappropriate. Consequently, the current National Director of Coaching agreed to be involved, as did his predecessor. It was felt that these people, with their respective cumulative coaching experience and professional association with the nation's top coaches, would provide an appropriate coach's perspective on refereeing performance criteria.

Secondly, representatives from the Australian Coaching Council (ACC) were not identical in terms of their respective professional roles. The ACC incorporates into its structure the National Officiating Program, however, the program includes only one staff member (the National Officiating Director). With the practical importance of this person's contribution to the BARS process recognised – plus the need to maintain a balance of expertise across each panel – a second representative from the ACC needed to be identified. Accordingly, the National Officiating Director nominated a person of similar understandings and knowledge of officiating issues to himself for inclusion on the panel.

Thirdly, academics on each panel were selected from departments of human movement and physical education at Sydney metropolitan universities. Both academics had extensive experience in sport as players and coaches, and coordinate units in sports education at their respective universities. Additionally, their inclusion followed panel membership specified in other sports-related BARS research (Anshel & Webb, 1991).

Following the identification of potential panel members, an indication of support was required from each expert before the constitution of panel membership was finalised. Consequently, all prospective panel members were contacted by telephone to gauge their interest in participating in the BARS meeting. Each was given a broad overview of the purpose for the meeting and the general procedures to be followed for the day. When verbal agreement from panel members was obtained, they were sent correspondence confirming their participation and providing administrative details for the meeting (see Appendix 2). Additional verbal confirmation was obtained from each panel member in the week prior to the meeting.

Overall, there were clear logistical problems inherent in bringing a specific, yet large, group of people together for one particular day. The tight focus of this investigation compounded this issue further, as the available sample for drawing appropriate elite expertise was small. Aside from this concern, panel membership was diverse with respect to the soccer perspective, and maintained a degree of expertise which is consistent with BARS methodology. The net result of these points is that the panel composition was suitable and appropriate given the character of this investigation.

## Identification of Support Personnel

Additional support staff were required to provide assistance on the day of the expert-panel meeting. The primary role of support staff was to maximise the input and utility of each panel member, plus provide assistance in compiling the final list of performance dimensions and competencies. Two types of support personnel were identified, namely, panel leaders (one per panel, as recommended by Anshel, 1995), and one process-observer (as recommended by Cyrs, 1979).

The role of the panel leader was central to the success of the meeting, as each was charged with the responsibility of ensuring the panel in their charge meets its objective for the day, i.e., a consensus on essential performance dimensions and competencies needed to officiate at the elite level of Australian soccer. To meet this objective, panel-leaders were required to undertake specific tasks (based on guidelines provided by Anshel, 1995; Cyrs, 1979) which included:

- clarifying any misunderstandings in terms of operational definitions;
- ensuring that views are solicited from all panel members;
- remaining 'neutral' in terms of the area of competence under discussion;
- ensuring that the panel discussion remains 'on-track' and that the panel does not digress into areas of deep philosophical discussion; and,
- paraphrasing and summarising statements, while also asking for clarification of language and meaning.

Due to the extensive interactive and facilitative role assigned to panel leaders, it was recommended that they have past experience in group management, and possess good communication skills (Anshel, 1995). In addressing this recommendation, panel-leaders for the present study were selected on the basis of three criteria, namely, their experience with BARS methodology; deference to their current professional roles (academics in physical education and sports studies); and, their professional backgrounds (trained educators in diverse applied settings, including child and adult education). Two applicable people were identified, and they agreed to undertake panel-leadership roles.

The role of process observer also required familiarity with BARS methodology. But unlike the diverse requirements of panel leaders, a more singular focus was needed. Specifically, the process observer was required to ensure a common approach was taken by panel leaders. This is achieved by observing each panel on a regular basis during the day (Cyrs, 1979). To facilitate this requirement, two procedures were adopted. Firstly, the process observer was

not 'attached' to any panel, and was free to visit each panel at random during the day. Secondly, the process observer met with the two panel leaders during the scheduled morning tea and lunch breaks to ensure consistent approaches, and to discuss any confounding issues that may have arisen during the course of panel discussions. Potential candidates, who were capable of meeting these requirements, were identified and approached, concerning their availability and willingness to be involved in the present study. One potential candidate declared their availability for this role.

### **Summary**

Although the procedures detailed in this section were numerous and logistically complex, the benefits of bringing together a diverse range of soccer-related experts exceeded any inconveniences. As a collective cohort, the breadth of soccer knowledge and understandings the cohort brought to the BARS process was substantial. Moreover, the diversity of the group ensured the referees role was examined from numerous perspectives. This reduced the potential of relevant performance criteria being overlooked (as may have been the case had one single group of stakeholders only been assigned to BARS panels).

Following the clarification and establishment of support personnel and panel membership, the expert-panel meeting was ready to proceed. This meeting is described in depth in the ensuing section, followed subsequently by a detailed review of the of the meeting's results.

### **MEETING OF EXPERTS – PERFORMANCE CRITERIA IDENTIFICATION**

This section describes the process undertaken to develop specific performance criteria of elite soccer referees in Australia. Initial discussion outlines procedures of a day-long meeting of soccer experts (as described in the previous section) to develop constituent performance dimensions and competencies. Particularly, discussion outlines the steps followed by each panel during the course of the meeting, and the subsequent final compilation of performance criteria.

In undertaking this aspect of the present study, the procedures followed an amalgam of techniques used to identify performance criteria in other sport-related roles (Anshel, 1995; Anshel & Webb, 1991; Anshel & Webb, 1989). For this study, procedures were broken into three stages. In Stage 1 (one-day duration), expert panels met to develop the requisite performance dimensions and competencies. Stage 2 incorporated a cross-referencing technique, where a composite list of performance dimensions and competencies was formulated from the initial work of the two expert panels. This stage, completed within days of Stage 1, was undertaken by panel leaders and the process-observer. Stage 3 acted as a reliability check, with panel members asked to verify the accuracy of the composite list developed in Stage 2. This final stage took approximately three-to-four weeks to complete.

## Stage 1 – Panel Meeting

The one-day meeting set aside to identify relevant performance dimensions and competencies was broken into four distinct parts, namely: meeting organisation; generation of action verbs and phrases; development of competencies; and, development of performance dimensions. All four parts are integral to the BARS process. A timeline outlining the implementation of these parts during the meeting is contained in Appendix 3.

To address organisational matters, all experts met together initially for the purpose of introducing panel leaders, the process observer, and other experts. The nature of individual roles and responsibilities were clarified at this point. To aid clarification, panel members were issued with handouts that outlined the aim of the meeting, various definitions they may require, and examples of similar work that had been completed to identify performance criteria of sports coaches (see Appendix 3). Time was allocated for any questions of the investigator concerning procedures for the day and the general thrust of the research.

Following the conclusion of organisational matters, experts were separated into two panels. In accordance with recommendations proposed by Anshel (1995), care was taken not to place individuals who had direct supervisory responsibilities over another person on the same panel. This measure is necessary to promote uninhibited discussion, and to minimise bias or intimidation during panel discussions (Anshel, 1995). From this point, each panel was required to work independently from each other. Responsibility for initiating performance dimensions and competencies now rested with each panel and their respective panel leader.

The first task allocated to each panel was to generate action verbs and phrases which described *essential* attributes for officiating soccer at the elite level. This involved each panel member working independently to develop their own list of action-verbs. The panel then came together to share their ideas and a composite list of verbs was subsequently recorded. Examples of action-verbs from this procedure included 'signal,' 'organise,' 'instruct,' 'react' and 'move.' From this list, panel members again worked independently to generate action-phrases, e.g., 'react quickly.' A composite list of action phrases was compiled and displayed for all panel members to review. During this process, no verb or phrase was rejected, thus ensuring as many actions as possible were listed for consideration by all panel members.

The second step of the meeting involved panel members working separately to write specific competencies. This task was accomplished by transforming each of the relevant verbs and action phrases into a statement that was either observable or measurable, e.g., 'communicates decisions with clear hand signals.' Consensus between panel members was obtained before the panel progressed to the next step.

The third and final activity for panel members was to organise competencies into what they perceived as related or homogenous categories, e.g., 'communication.' Once the categorisation was completed, panel members returned to their group to define each category – also referred to as performance dimensions – as statements (e.g., 'demonstrates effective communication skills'). Performance dimensions are designed to be reflective of the constituent set of competencies.

At the conclusion of this third activity, there was a final gathering of all panel members. Here they were thanked for their participation and contribution to the research. In total, Stage 1 took approximately six hours.

### **Stage 2 – Amalgamation**

Following the completion of Stage 1, two distinctive lists of performance dimensions and competencies had been developed. These needed to be combined into one list, without losing meaning or intent from either list. To achieve this task, the process observer and each panel leader met, and amalgamated both lists into one combined draft list of performance dimensions and competencies. This procedure, also referred to as the *retranslation*<sup>1</sup> stage (Anshel, 1995), allowed similar performance dimensions and competencies, identified by each panel, to be integrated.

As an additional safeguard to ensure that all performance criteria were included, dimensions and competencies that were specific to each panel were retained. This guaranteed that the role under investigation was comprehensively covered, and no competency or skill was excluded.

### **Stage 3 – Panel Feedback**

In accordance with operations outlined in the literature (Anshel, 1995; Anshel & Webb, 1991; Cyrs, 1979; Cyrs et al., 1976), the draft list of performance dimensions and competencies compiled from the amalgamation process was sent to all panel members for comment (see Appendix 4). This procedure ensured that the intent and content of performance dimensions and competencies had not been lost in the amalgamation process. Panel members were asked to comment on any editorial concerns (e.g., rewording) and to consider a number of questions relating to the draft list. These included:

---

<sup>1</sup> This term is borrowed from traditional BARS literature. However, the interpretation of 'retranslation' by Anshel (1995) is not in accordance with retranslation descriptions provided by Smith and Kendall (1963) and others (see Chapter 2). The redefinition of the 'retranslation' phase by Anshel is examined further in Chapter 4.

1. Are there any performance dimensions that you think need to be added, deleted or combined?
2. Do the competencies reflect accurately the performance dimension with which they are associated?
3. Are there any competencies that you think are ambiguous?
4. Are there any competencies that you think are superfluous?
5. Are there any competencies that you think need to have specific examples included (for clarification purposes)?

Minor alterations were suggested and are outlined in Appendix 5. All proposed changes were submitted to the process observer and panel leaders for review. Judgements relating to possible item alteration were based on two criteria, namely, is the alteration supported by other respondents, and does the proposed alteration retain the meaning and intent of the item. Consensus between support staff with respect to these criteria was necessary before any alteration was accepted into the final list. Relevant alterations were made, and a final list was agreed (this list is discussed in detail in the following section).

### **Evaluation of the Preliminary Study Design**

Although methodological design implications of the preliminary study are addressed in the following chapter (specifically, see *Evaluation of Research Design*), a number of qualitative aspects, associated with the preliminary study, are worthy of examination here. As noted by Hakim (1987), the qualitative research is commonly used for exploratory studies, which in turn can form the basis of additional studies that are more structured or quantitative. In particular, Krathwohl (1993) recommended qualitative designs when specific circumstances prevail. The most pertinent of these circumstances, as they relate to the present study are listed below. These are accompanied by explanations which reflect the circumstances specific to this investigation, and provide criteria to assess the validity of the preliminary study's design. These criteria included:

1. *research is lacking in an area and must emphasise discovery.* As outlined in Chapter 1, empirical research relating to the requirements of elite soccer refereeing performance is scant. Consequently, the present investigation, through the preliminary study, set out specifically to 'discover' which performance criteria were essential for the role of elite soccer refereeing.

2. *there is no valid and trusted measures of the phenomenon of interest.* At this point in time, detailed instruments that describe or measure the performance of officials have not been developed.
3. *research progress in the area has plateaued, and fresh perspectives are needed.* The previously discussed studies into officiating by Anshel (1995) and Anshel and Webb (1991) did not target a specific officiating level within each sport, i.e., elite, intermediate, beginner, etc. Their approaches subliminally inferred that performance criteria, and the underlying importance of specific criteria, were consistent across all competitive levels of sport. Such assumptions contradicted the highly focused nature of BARS generated data as described by Shirom (1988). Consequently, it is necessary to move officiating research another step forward and focus attention on performance requirements within a specific level of sporting performance.
4. *there is a need to explore a phenomenon in depth and lean about it.* As the requirements of the referees role are diverse, with little agreement about the specifics of effective performance, it is difficult to detail accurately and completely the nature of the referees role. This problem has not escaped the attention of Clegg and Thompson (1993), who surmised that if 100 officiating experts were to submit their own list of the essential qualities of the master official, one might expect 100 different listings. A lack of empirical investigation has not helped in this matter. New approaches are needed to explore the referees role and related issues in greater accuracy and depth.

These four circumstances supported strongly the qualitative emphasis given to the initial stage of the present study. Moreover, the foundational data-platform generated from the preliminary study is a key feature of qualitative design (Hakim, 1987). More particularly, the preliminary study borrowed qualitative elements normally associated with ethnography – a specific paradigm of qualitative research. Techniques associated with this paradigm allow descriptions of a group or culture, with “patterns of human thought and behavior [being] the focus of inquiry” (Fetterman, 1989, p. 11). Ethnographical aspects of the investigations research design are addressed in the following chapter (specifically, see *Evaluation of Research Design*), however, one element of ethnographic research is relevant to discussion here.

This issue focuses on the researcher perceiving the ‘big picture’ of the situation rather than focusing on specific elements within the situation (Borg & Gall, 1989, p. 389). The generation of specific refereeing competencies, which in turn were representative of somewhat discrete refereeing performance dimensions, seemed to contradict this ‘big picture’ concept. However, the BARS process required an holistic perspective of the referees’ role to be embraced. The apparent dichotomy occurred because the identified competencies are highly specific, yet the retention of performance dimensions and competencies which are

common *and* unique to each expert group were retained. Consequently, the resultant performance dimensions and competencies formed a comprehensive overview of the role of the referee.

### Summary

The procedures undertaken for performance criteria identification were predicated on those used previously by Anshel and Webb (1991) and Anshel (1995) for identifying officiating performance criteria in Touch football and basketball, respectively. Additionally, the approaches and procedures undertaken to identify performance criteria is consistent with qualitative methodology. In the specific context of present study, the use of two panels of experts, from diverse soccer and officiating roles, ensured the widest possible perspectives were brought to the competency identification process. Moreover, the hybrid BARS procedures, as described in this section, ensured performance criteria were developed in a systematic and coordinated manner, thus minimising the potential for the inclusion of criteria which lacked relevance to the role under investigation.

## RESULTS AND DISCUSSION – PERFORMANCE DIMENSIONS & COMPETENCIES

The BARS process outlined in the previous section produced a diverse range of soccer refereeing performance criteria. This section details the outcomes of that process. Specifically, the preliminary study identified 38 competencies, spread throughout six performance dimensions. Each of these dimensions and competencies were identified as essential refereeing performance criteria for elite soccer referees in Australia. To clarify the presentation and discussion of these results, performance dimensions are presented in bolded text, with related competencies listed beneath each dimension. General discussion, pertinent to each dimension, is then provided.

### **1. Displays technical skills and knowledge associated with soccer officiating**

- understands and interprets the Laws of Soccer correctly
- applies the Laws of Soccer
- understands how the game is played (e.g., tactics and strategy, analyse patterns of play)
- anticipates players actions and reactions (e.g., retaliation to over physical play)
- keeps a complete record of the game (e.g., bookings, goals, substitutions)
- works as a team with assistant referees
- observes incidents and decides on appropriate action (e.g., fouls, player acting free kick, booking, play-on)
- observes play from the best position (not too close to the ‘action’)

This performance dimension encompasses two aspects of referee performance, i.e., technical refereeing skills and refereeing knowledge. Although it can be debated that these two aspects

of performance can be viewed separately, the expert panels saw these areas very closely related. Further, during panel discussion, the experts argued that to separate these performance aspects had little practical foundation.

The eight competencies listed within the performance dimension tend to support this notion. In particular, the competencies are inclined to compliment each other. An example can be found in the two competencies, *observes play from the best position (not too close to the 'action')*, and, *understands and interprets the Laws of Soccer correctly*. These competencies represent, respectively, a discrete technical competency and a discrete knowledge competency. However, an obvious symbiotic relationship is evident. Clearly, there is significant practical value for the referee to be in the best position to view an incident. This enables referees to apply their understanding of the Laws to what they are viewing. Similarly, correct understanding and interpretation of the Laws is of minor consequence if the referee is not in position to view adequately game incidents. These findings demonstrate the discrete nature of competencies within the dimension, while, at the same time, reflect the inter-relationship between knowledge and technical skills for effective refereeing performance.

The competencies also provide clear evidence that the role of the referee is not exclusively task oriented. A number of competencies, including, *applies the Laws of Soccer*, and *understands how the game is played (e.g., tactics and strategy, analyse patterns of play)*, indicate a cognitive aspect to the referees role. Implicitly, these competencies aid in the successful performance of less complex competencies, yet are not evident necessarily through explicit occupational behaviours.

Lastly, the competency, *works as a team with assistant referees*, testifies the referee's role is not independent of other officiating peers. In particular, referees work in conjunction with assistant referees to officiate a game. If refereeing performance is to be successful, referees needs to rely on assistant-referees to provide quick and accurate information concerning pertinent decisions. Without such support, there is potential for referees to make decisions based on incomplete information, thus increasing the probability of decisions being incorrect.

## **2. Applies the Laws to maintain high levels of accurate decision making**

- applies the Laws consistently (within and between games)
- observes and analyses incidents and correctly interprets such incidents
- reacts quickly and effectively to incidents
- distinguishes between fair and foul play
- interprets and discriminates between the continuum of fouls
- encourages attacking play
- distinguishes between advantage and disadvantage
- manages conflict (quick decision making, number of free kicks – red and yellow cards)
- moves to obtain optimum positions (i.e., place with best view and close enough to react effectively)

This dimension is concerned primarily with the decision-making role of the referee. The role encompasses a number of elements, including: rule adjudication; rule interpretation; positional work; and, consistency of rule application. These elements are reflected in the number ( $n=9$ ) and diversity of competencies within this dimension.

Notably, the competencies cover both the mechanics *and* processes of decision-making, i.e., where to make decisions, what to adjudicate, and how to adjudicate. For example, *moves to obtain optimum positions (i.e., place with best view and close enough to react effectively)*, specifies that the referee needs to move to a position that is relatively close to the play, and affords the best possible view. The referee then needs to be able to make a decision quickly, i.e., *reacts quickly and effectively to incidents*, and ensure the decision is consistent with other decisions made during the game, i.e., *applies the Laws consistently (within and between games)*. By prescribing the competencies related to decision-making in this way, a clear picture emerges that decision-making is not limited to adjudicating incidents as they occur, but involves an ongoing process within, and between, games.

One competency that caused debate within the expert panels was *encourages attacking play*. Some panel members felt this competency to be the responsibility of coaches and players, and moreover, referees had little or no responsibility in this regard. However, the competency's inclusion was established through its link with the competency *distinguishes between advantage and disadvantage*. As this latter competency was viewed, without argument, as an essential refereeing competency, it was agreed by panel members that the correct application of advantage can promote attacking play. Given this close link between the two competencies, *encourages attacking play* was included in the final list.

More generally, competencies such as those discussed in the previous paragraph show many competencies are predicated on referee interpretation (especially interpretation of Laws, and Law application to game incidents). Other examples include, *distinguishes between fair and foul play*, and *interprets and discriminates between the continuum of fouls*. Inclusion of competencies such as these demonstrate the difficulty in describing occupational performance in behavioural terms alone. Although it can be argued the consequences of these decision-making competencies may be observed through referee behaviour, clearly, decision-making is a cognitive skill. As such, their inclusion is acceptable, and necessary, given the penchant of the hybrid BARS process for describing occupational performance.

### **3. Demonstrates effective communication skills**

- effectively uses the whistle (e.g., volume, tone, timing, length, player reaction)
- communicates decisions with clear signals
- undertakes report writing and record keeping to maintain a complete record of the game (e.g., send off reports, administrative reports)
- communicates (verbally and non-verbally) with players on and off the field
- manages disputes between opposing players

- communicates confidently with assistant referees (e.g., before the game on expectations, non-verbal corresponds with the verbal and body language)
- communicates confidently with coaches (e.g., handle yourself appropriately, conflict resolution, positive attitude)
- communicates with inspectors (e.g., post match discussions, self reflection, constructive criticism)

This performance dimension is designated specifically to the communication of referees with soccer-related peers. These peers include other officials (assistant referees), players, coaches, and referee inspectors. Similarly, the forms of communication employed by the referee, i.e., speech, hand signals, whistling, and report writing, vary also. Such diversity in the communication role necessitated identification of eight communication-specific competencies.

The forms of communication strategies described by the competencies relate to specific circumstances. However, they can be divided into two aspects, i.e., player/coach communication, and communication with officiating peers. For communication with players and coaches, the primary reason the referee communicates with these groups is to relay decisions. Within this context, hand signals and whistling predominate. This fact notwithstanding, in instances of conflict resolution, verbal communication would also be incorporated.

Communication with officiating peers is similarly varied. This is evident particularly for communication between referees and assistant-referees. As communication between officials is a two-way process (and across relatively vast distances at times), communication is conducted through a variety of non-verbal signals (hand and flag) and verbal communication. Lastly, communication with referee inspectors is less complex, i.e., verbal and non-verbal (report writing), as communication is concerned primarily with assessing the performance of referees.

Overlaying these aspects of communication is a requirement for communication competencies to be employed on and off the field of play. Competencies verify this is the case when referees communicate with all stakeholders (the only exception being referee inspectors, i.e., off-field only). Clearly, this communication requirement, and those noted previously, indicate the referee needs to develop a comprehensive range of communication skills, and be prepared to use them in a variety of contexts.

#### **4. Maintains standards and codes of behaviour**

- applies the Laws and sanctions (e.g., did the referee enforce a code of behaviour?)
- displays positive behaviours and feedback
- manages disputes between players/coaches and officials
- monitors player behaviour
- management of all aspects of the game (e.g., match control/players)

This dimension relates to the behaviour of the referee *and* players. The dimension applies specifically to on-field behaviour, and, as a consequence, tends to focus predominantly on player management. It is noted also that the number of competencies comprising this performance dimension ( $n=5$ ) is notably less than for the dimensions discussed previously. Reasons for this circumstance are not obvious from the data or the BARS process. However, the variation in competency numbers for respective performance dimensions is consistent with other hybrid BARS studies.

As mentioned, the focus of this dimension is toward player management (four of the five competencies can be interpreted in this manner). This result can be explained through two game-related factors. Firstly, the combative nature of soccer means that players will, from time-to-time, engage in behaviour which is overly vigorous and can endanger the well-being of other players. Secondly, the elite standard of play ensures players will push the legal limits of physical play to obtain their desired objectives. In combination, these two factors can create competitive environments where player behaviour is not in accordance with the Laws of soccer, thus necessitating the development of refereeing competencies to address such play behaviour.

#### **5. Displays sound self management and personal attributes**

- prepares well in advance of match (revision of the laws, time of arrival at ground – on time, presentation on arrival, kit prepared)
- undertakes mental preparation for the match (e.g., visualisation)
- manages personal anxiety
- undertakes post match expectations (e.g., talk to coaches, attend post game functions)

This performance dimension relates to the personal attributes referees demonstrate away from the competitive match context (i.e., in a pre-game and post-game context). The inclusion of performance criteria undertaken ‘off-field’ raised considerable debate within each expert panel, and represents a marked change from the four previous dimensions (where an on-field foci predominated). However, as competencies which comprise this dimension were viewed as essential to effective performance, their inclusion met with consensual agreement from the expert panels.

In total, four competencies were classified into this performance dimension. Specifically, the first three competencies are pre-match considerations, and the last a post-match consideration. With respect to the pre-game competencies, the first competency can be viewed in the context of ‘professional preparation.’ By undertaking the tasks described in this competency, the referee is able to project a professional approach to their work that is expected of participants in elite sport. The next two competencies relate to the referee’s psychological preparation. These are similar in nature to those undertaken by elite athletes,

and are recognised as a fundamental prerequisite for elite performance. Inasmuch, these three competencies were seen to underpin successful on-field performance.

The final competency from this group, *undertakes post match expectations (e.g., talk to coaches, attend post game functions)*, is interesting as, at face value, it has the least effect on on-field performance than any competency within this, or the other five, performance dimensions. The execution of this competency occurs after the completion of a match, and the proficiency with which it is undertaken is of no consequence to the referees' control of the game. However, the competency was, nonetheless, seen to be essential to referee performance in a broader context. Furthermore, it is consistent with competencies in this performance dimension (i.e., not undertaken on the field of play), and adds further evidence to demonstrate the referees role is not confined to on-field performance.

## **6. Engages in appropriate personal health and safety strategies**

- maintains required levels of fitness (e.g., fitness benchmarks/standards, hydration, warm up/cool down)
- maintains appropriate levels of personal health (e.g., correct diet, weight control)
- undertakes appropriate risk management procedures (e.g., knows legal responsibilities, environmental changes to conditions)
- maintains concentration during the game

This performance dimension links aspects of health and safety that encompass referee *and* player well-being. The dimension is underpinned by the move to a more professional approach to elite sport. In generations previous to this, the health and fitness of the referee may have been of minor concern. However, the improved fitness of players, and the corresponding increase in game speed, has increased attention toward this aspect of referee performance. Similarly, the professional environment has seen the legal responsibilities of referees brought into sharper focus. Consequently, referees are not only responsible for administering the Laws of soccer, but also for creating a safe playing environment for the players.

Four competencies were classified into this dimension. The first two competencies listed are aimed specifically at the health and fitness of the referee. These competencies imply referees are required to not only look after their fitness, but to engage in practices which promote good health. As a matter-of-course, this necessitates that these competencies are not undertaken in the game-context. More likely, they are undertaken in the week/s before and after a game. However, their affect on game-performance should not be ignored, as the link of fitness and health to accurate decision-making is well established (see Chapter 1).

Of the remaining two competencies in this group, the third competency, *undertakes appropriate risk management procedures (e.g., knows legal responsibilities, environmental changes to conditions)*, is aimed primarily toward player safety. Similarly, the focus of this

dimension is intended to shield the referee from possible legal action arising from player injury. As described in Chapter 1, the legal responsibilities of referees for player safety are now paramount, and neglect of these responsibilities can result in the referee being held legally liable for player injury.

The final competency, *maintains concentration during the game*, may seem, at face value, to be a curious inclusion. In particular, the competency appears to be unrelated to the other three. However, the expert panel that provided this competency argued strongly that safety is of such fundamental concern, referee concentration on this aspect of performance is crucial. Accordingly, the competency was placed within this performance dimension. More generally, the inclusion of this competency is indicative of the hybrid BARS process, where competencies are not discarded unless there is consensus from the expert panel.

### Summary

As detailed in Chapter 2, hybrid BARS is premised on behavioural aspects of performance. However, the procedure allows provision for the identification of other performance attributes including knowledge, understanding, and attitudes. This integrated approach to competency development provides a richer understanding of occupational requirements (see Hager & Beckett, 1995; Preston & Walker, 1993). The resultant competencies identified in this preliminary study reflected this approach. The majority of competencies were based on behavioural skills and tasks, however, a number of competencies were founded on the performance attributes of understanding and knowledge, e.g., *understand and interpret the Laws of Soccer correctly*, and, *distinguish between advantage and disadvantage*. Although such attributes may be difficult to assess in performance terms, care was taken to ensure all competencies were either observable or measurable. Subsequently, results have provided competencies, categorised into performance dimensions, that took into account the diversity and complexity of refereeing performance, and were clear enough to render them appropriate for assessment purposes.

### FINDINGS, IMPLICATIONS, AND CONCLUSIONS

In the previous sections, discussion has described three aspects of the preliminary study, i.e., the context of the study, the procedures followed to identify performance criteria of elite soccer referees, and, resultant performance dimensions and competencies. This section examines the implications of those procedures and results, and answers the research questions outlined in section one of this chapter.

## Findings and Implications

This chapter addressed two research questions from Theme 1, namely:

- 1.1 What performance dimensions and competencies, identified using hybrid BARS development procedures, are required for elite soccer referees?
- 1.2 Are performance dimensions and competencies reflective of officiating performance criteria in other sports?

With respect to Question 1.1, the data generated through the hybrid BARS process yielded 38 competencies across six performance dimensions. The range of competencies was diverse, and encompassed specific 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). The hybrid BARS data have presented performance criteria specific to the soccer context. In this regard, the criteria are more narrowly focused and relevant to soccer. As such, the criteria present a unique perspective on the performance requirements of elite soccer referees.

Three specific points in relation to the data are worthy of comment. Firstly, the first three performance dimensions described in the list accounted for 26 of the 38 competencies. This represents approximately two-thirds (68%) of the total number of competencies. The emphasis on these three performance dimensions is not surprising, especially given the core roles these dimensions play in refereeing performance (Australian Coaching Council, 1996; National Officiating Program, 1997; Weinberg & Richardson, 1990). Furthermore, the number of competencies within these three performance dimensions reflect the breadth of competencies referees are required to embrace for effective performance. For example, with respect to communication, referees are expected to communicate through a variety of mediums (whistle, hand signals, text, verbal), and to a range of stakeholders (players, coaches, referee inspectors, assistant referees). Therefore, it would not be unreasonable to assume the more fundamental a performance dimension is to effective performance, the greater the number, and diversity, of related competencies that are required for elite refereeing performance.

Secondly, it is of interest to note that, in two instances, similar competencies were included in more than one performance dimension, e.g., *observes play from the best position*, and, *moves to obtain optimum positions (i.e., place with best view and close enough to react effectively)*. Despite their similarity, the subtle wording difference between the two competencies portrays marked variation in performance requirements. The former is implicitly related to where the referee observes play, while the latter is related to the mobility of the referee to obtain that position. Such subtle differences, while seemingly negligible to

the lay-person, demonstrated the value in having competencies derived by experts who are familiar with the role under investigation.

Thirdly, the inclusion of similar competencies confirmed the dynamic and complex nature of referee performance. This phenomena was previously brought into sharp focus in traditional BARS literature, where problems arose in determining wholly discrete occupational performance dimensions (Burnaska & Hollmann, 1974; Jacobs et al., 1980; Schwab et al., 1975). This was due primarily to inter-relationships existing between many occupational competencies, yet, nonetheless, placed into separate performance dimensions. However, as noted by Schwab et al. (1975), the determination of conceptually distinct performance dimensions is an unrealistic expectation given the nature of work. The results of this study tended to affirm is notion.

In regard to the second research question, the performance dimensions and competencies detailed are generally consistent with performance requirements described in Chapter 1, i.e., communication, rule interpretation, interaction with players and coaches, and litigation. Moreover, the data reflected performance dimensions and competencies which underpin officiating across all sports, e.g., communication and decision-making. This notion is conceptualised in the development of officiating 'general principles' courses by the Australian Sports Commission (National Officiating Program, 1997).

Despite the consistencies of the results of this study with the literature, it should be noted also that the number of performance dimensions and competencies developed in this investigation varied to those in earlier officiating studies on basketball (Anshel, 1995) referees and Touch referees (Anshel & Webb, 1991). These variations should be expected. It would be inappropriate to assume that officiating across different sports would require exactly the same performance requirements. The specific and varying characteristics of officiating in any given sport, or at any given level of a sport, should, arguably, result in different skill requirements of officials. For this reason, each sport should guard against directly equating the roles of one group of officials with that of another. The results from the analysis in this study confirm that specific sports need to analyse the performance requirements of their officials carefully and systematically. Further, there should be an expectation that the competencies identified would be different for other sports.

### **Conclusion**

The application of hybrid BARS methodology proved to be successful in the generation of competencies and skills required of elite soccer referees in Australia. The process, which utilised a cross-section of soccer experts, was able to 'tap' attitudes and perceptions of soccer stakeholders that may normally be excluded through alternative processes. This ensured the referees' role was examined thoroughly from a variety of perspectives, and minimised the

possibility of pertinent or relevant performance requirements being omitted. As a result, the identified performance dimensions and competencies encompassed a broad range of skills, actions, roles, and responsibilities the referee undertakes in fulfilling their duties. The breadth and range of requirements were not surprising, and reflected the complex nature of officiating alluded to in the literature (Anshel, 1995; Anshel & Webb, 1991; Australian Coaching Council, 1996; Dickson & Webb, 1998; Evans, 1994; National Officiating Program, 1997; Weinberg & Richardson, 1990).

The identification and generation of these performance dimensions and competencies allows for additional investigation into other officiating issues, i.e., the *importance* of individual refereeing competencies, and the *preparedness* with which competencies are undertaken. To meet this end, a second investigation was initiated. The next chapter outlines relevant aspects of the research design, with a particular focus on the development of a research instrument predicated on the referee performance dimensions and competencies discussed in this chapter.

## CHAPTER FOUR

# PRINCIPAL STUDY RESEARCH DESIGN AND METHODOLOGY

### Introduction

The preliminary study, described in Chapter 3, answered research questions from Theme 1. In doing so, the resulting performance dimensions and competencies allow the remaining research themes to be investigated. This latter investigation comprises the 'principal study,' and procedures undertaken for the study are reviewed in this chapter.

In particular, this chapter provides detailed descriptions of the research design and methodology implemented in this investigation, with initial discussion focused on the context in which the present study was set. Subjects comprising the research sample are also defined in this initial section. Issues relating to constraints imposed on the research, the nature and integrity of the research design, and the techniques employed to analyse data, are explored in later sections. In the last section, a detailed explanation of the Rasch Rating Scale model, an innovative technique for a study of this nature, is undertaken.

Although discussion focuses primarily on the principal study, aspects of the preliminary study are included. This approach provides a holistic perspective on the entire investigation, and illustrates the close relationship between the two studies. Specifically, sections relating to *Planning for Research, Research Design and Procedures, Instrumentation, and Evaluation of Research Design* comprise an integrated approach for discussion.

### CONTEXT AND PARTICIPANTS

This investigation is focused predominantly on exploring issues associated with elite soccer referees. Accordingly, analysis of the research context takes two perspectives, namely, the structure of the organisation which administers Australian soccer, and the character of Australian soccer at the highest level.

#### Context

At the macro level, soccer in Australia is governed by the Australian Soccer Federation, which now operates under the corporate name of Soccer Australia. It is a non-profit

organisation, and during the 1996-97 season, had a financial turnover in excess of \$7.2 million (Soccer Australia, 1997a). These funds are spent on the many tiers of Australian soccer, which encompassed over 500,000 participants across the country, ranging from junior to full international status (Soccer Australia, 1997a).

As a member of the Federation Internationale de Football Association (FIFA), Soccer Australia is responsible for the administration of soccer according to the FIFA Statutes, Regulations and Decisions. Therefore, the Federation agrees to cooperate in all matters relating to the organisation of international competitions and soccer in general (Soccer Australia, 1997b). As an adjunct to this undertaking, Soccer Australia is also responsible for marketing and promoting soccer throughout Australia. This involves: supplying national teams for international soccer competitions, such as the World Cup and Olympic Games; the administration of national competitions (at youth and senior levels); the implementation of youth development programs; and, the provision of refereeing and coaching programs.

Of all domestic competitions administered by Soccer Australia, the one with the highest profile is the national league. First established in 1977, the national league is the premier domestic competition in Australia, and is now known as the Ericsson Cup (Soccer Australia, 1997a). The Cup is regulated by Soccer Australia through a number of sub-committees. However, all policy and disciplinary decisions in relation to this competition are vested in the Executive Committee of Soccer Australia (Soccer Australia, 1997b). Soccer Australia also exercises some control over the players and coaches who participate in the Ericsson Cup, as these parties must abide by the provisions of the Articles of Association of the Australian Soccer Federation (Soccer Australia, 1997b).

As with other football codes in Australia, elite soccer is moving to full-time professional status, particularly among many players and coaches. These two groups are employed by respective Ericsson Cup clubs on either a full-time or part-time basis. Players are normally selected by clubs for their ability to be competitive at the elite level, while the coach's appointment is based on his/her potential to win matches using the players at their disposal. To a lesser extent, coaches are also required to engage in player development. Due to the fickle nature and high demands of professional sport, if match results are not to club expectations, it is not uncommon for coaches to be dismissed before the completion of a full season, and players transferred to other national league, or lower division, clubs.

Aside from the contractual obligations players and coaches have to their employing club, each are bound by the provisions of the Articles of Association of the Australian Soccer Federation (Soccer Australia, 1997b). Players are particularly affected by this obligation, as they are expected to make themselves available for representative matches from time-to-time as approved by Soccer Australia (Soccer Australia, 1997b).

Control of match-play between Ericsson Cup clubs is deferred to match officials (referees and assistant referees), who are employed and appointed by Soccer Australia. The selection of referees and assistant referees to officiate on a week-to-week basis is made from the national league 'panel.' The panel, comprising 13 referees and 36 assistant referees, provides a pool of officials to control all games. Membership of this panel results usually from recommendations by respective State Referees Commissions to the National Referees Director (pers. comm. Gary Power, 7 May 1997). To gain recommendation, four essential criteria have to be met (Power, 1997b). These are:

1. to be one of the top referees in their state;
2. to pass the FIFA Fitness test;
3. to be capable of continual improvement; and,
4. to provide a medical clearance from a registered medical practitioner.

Although officials receive payment from Soccer Australia for carrying out their officiating duties, they are not considered fulltime employees, and many are employed outside the soccer industry.

### **Definitions of Soccer Stakeholders**

The research questions and hypotheses addressed in the present study are designed to elicit opinion on refereeing performance criteria from a range of stakeholders involved in the Ericsson Cup competition. Accordingly, this sub-section describes stakeholders (i.e., participants) associated with this investigation, and clarifies the soccer-related roles assumed by each stakeholder group. Notably, all stakeholders participate at the elite level of soccer, and are of male gender. The stakeholders are referred to as players, coaches, referees, assistant referees and referee inspectors. Although these groups were alluded to in the previous sub-section, ensuing discussion describes in detail the specific roles assumed by each group.

#### **Players**

Players constitute those people who are registered with an Ericsson Cup club and *played* in the Ericsson Cup competition during the 1996-97 season. This discriminates Ericsson Cup players from other players who constitute larger Ericsson Cup squads, but for various reasons, take no part in Ericsson Cup matches (this latter group of players normally compete in the National Youth League, a secondary competition to the Ericsson Cup).

## Coaches

Coaches include those people appointed by Ericsson Cup clubs to prepare players for Ericsson Cup competition on a week-to-week basis. Preparation normally includes strategic and tactical development, skill practice, fitness development and maintenance, and psychological preparation (National Coaching Director, Selby, J. 1997, pers. comm. 13 June). In this study, coaches were defined as Ericsson Cup head coaches and assistant coaches, who were currently serving at the time of data collection<sup>1</sup>.

## Referees, assistant referees and referee inspectors

This group of people are those appointed by Soccer Australia to fulfil the respective roles of referee, assistant referee, and referee inspector, for the 1996-97 season. These three sub-groups comprise discrete membership, with no overlap occurring between groups. For example, if an individual is appointed as a referee for the season, he cannot assume any assistant referee or referee inspector duties for the season. The specific roles assumed by each group is reviewed below.

### Referees and Assistant Referees

Matches between Ericsson Cup clubs are governed by one referee and two assistant referees (collectively, these two groups are referred to as 'officials') appointed by Soccer Australia from the national league panel. Specifically, the referee "has full authority to enforce the Laws of the Game in connection with the match to which he is appointed" (Soccer Australia, 1998, p. 16). Decisions of the referee are final, however, the referee may change a decision based on *advice* only from an assistant referee (Soccer Australia, 1998). For the purposes of this study, the term referee is used exclusively to denote these people described in the above definition.

Two assistant referees (commonly referred to as 'linesmen') are appointed to aid the referee in controlling a match (Soccer Australia, 1998). Their specific duties are varied, but they are required basically to indicate: if the ball is out of play and which team is responsible to restart play; if a player is in an off-side position; when a team requests a player substitution; and, when incidents or misconduct occur which are out of the view of the referee (Soccer Australia, 1998).

---

<sup>1</sup> Not all clubs employ assistant coaches. Assistant coaches were surveyed at Ericsson Cup clubs where their employment was indicated by respective head coaches. For analysis purposes, head and assistant coaches were classified into one group.

### Referee Inspectors

Soccer Australia monitors the performance of its officials through the use of referee inspectors. They are generally senior referees who have diversified their officiating career into the inspecting role. Specifically, they are employed by Soccer Australia (on a part-time basis) to assess officiating performance at all Ericsson Cup matches. In doing so, they are required to furnish detailed performance information to the referee, and Soccer Australia, through written and verbal feedback. It is common practice for only one inspector to be used per match.

### Summary

The dissimilar roles undertaken by respective stakeholders provide a unique dimension to the present study. Although, in one aspect, the stakeholders can be considered homogenous (i.e., they are all involved in soccer, and, more specifically, involved in elite soccer), they are nonetheless heterogeneous. In particular, players and coaches have the primary objective of winning games, and, in doing so, focus on strategies and tactics which meet this outcome. Any consequential 'domino effect' of not winning starts with the players. If players are not performing to expected standards, they run the chance of being dropped from the team. In turn, if poor results persist, the coach's employment may be terminated by the employing club. Such emphasis on game outcome is not evident in the roles of officials and inspectors. These people are more concerned with ensuring the game progresses in an orderly manner through the responsible administration of game rules. In doing so, they tend to focus on the numerous incidents that require adjudication, rather than one specific aspect of the game (e.g., the final score).

## PLANNING FOR RESEARCH

The following discussion examines challenges to the research, and, in doing so, provides planning solutions to meet such challenges. In particular, while planning the program of research, the investigator had to account for constraints imposed by the context of elite soccer in Australia. Specifically, two inter-related issues required careful consideration. The first was maximising data collection from participants who were spread across Australia, and the second, obtaining access to participants. As the two issues are inter-related, solutions to one issue is likely to impact on the other.

Firstly, the need to maximise data collection from disparate and diverse groups emerged as a significant challenge for the research. In illustrating this challenge, it should be recognised that subjects resided in all States and Territories of Australia with the exception of Tasmania and the Northern Territory (as these regions do not have clubs which participate in the Ericsson Cup). Consequently, the logistical demands placed on this research made data

collection problematical. Ostensibly, various solutions were required to maximise the quantity and quality of data collected (specific procedures are detailed later in this chapter – see *Research Design and Procedures*).

Further compounding this constraint were attitudes held by different stakeholders to administrative procedures. Specifically, advice from Soccer Australia (Power, G. 1996, pers. comm., 19 September) indicated that players are often neglectful in returning any form of written correspondence, questionnaires, or official documentation. As a balance to this, referees were said to be generally enthusiastic about initiatives and programs related to their soccer status (Power, G. 1996, pers. comm., 19 September). Subsequently, it was recognised that the implementation of various methods were necessary if data collection was to be maximised across each stakeholder group.

Access to each group of stakeholders, particularly Ericsson Cup players, also presented a significant challenge to the data collection process. The possibility of using match days was rejected, as only eleven players and substitute players would be available. Moreover, both players and coaches tend to be preoccupied with the upcoming match, and are inclined to take an adverse view of anything they perceive as ‘unnecessary distractions’ at this point. Immediate post-match periods were also deemed inappropriate, as match post-mortems usually occur at this time, and players are also highly fatigued from 90 minutes of play. Due to these factors, and the advice offered by Soccer Australia concerning player ambivalence to ‘administrivia,’ it was determined that the most effective and efficient way to access this sub-group was through the club training environment. During training sessions, full squads of players would generally be available, and coaches would be more willing to allocate time for data collection during this period. Accordingly, training sessions were seen as the most appropriate time to collect data from players in order to maximise sample size.

Similar issues concerning access to referees, assistant referees and referee inspectors did not arise. This was because their employer – Soccer Australia – was supportive of the present study and actively facilitated access to these groups.

A number of implications arose from the constraints noted in this section. Firstly, the need to collect data from players at their training venues, and across such a large geographical area, necessitated the employment of research assistants. This, in turn, required the recruitment and training of research assistants for this task (for specific details, see *Research Design and Procedures*). Secondly, data collection needed to be organised and arranged within the context of each club’s training schedules and commitments. This implication presented inherent problems for data collection. Training sessions, although standardised to a certain extent with regards to scheduling, can vary from week-to-week and session-to-session in terms of content and player participation. Thirdly, as a consequence of obtaining player data during training periods, the data collection needed to occur ‘in-season.’ Overall,

when planning to collect data from this group, flexibility in data collection procedures were paramount.

## RESEARCH DESIGN AND PROCEDURES

This section provides an overview of the research design (incorporating the preliminary study). Specifically, it details the procedures that were employed to investigate the research themes, questions and hypotheses identified in Chapter Two.

### Design of the Study

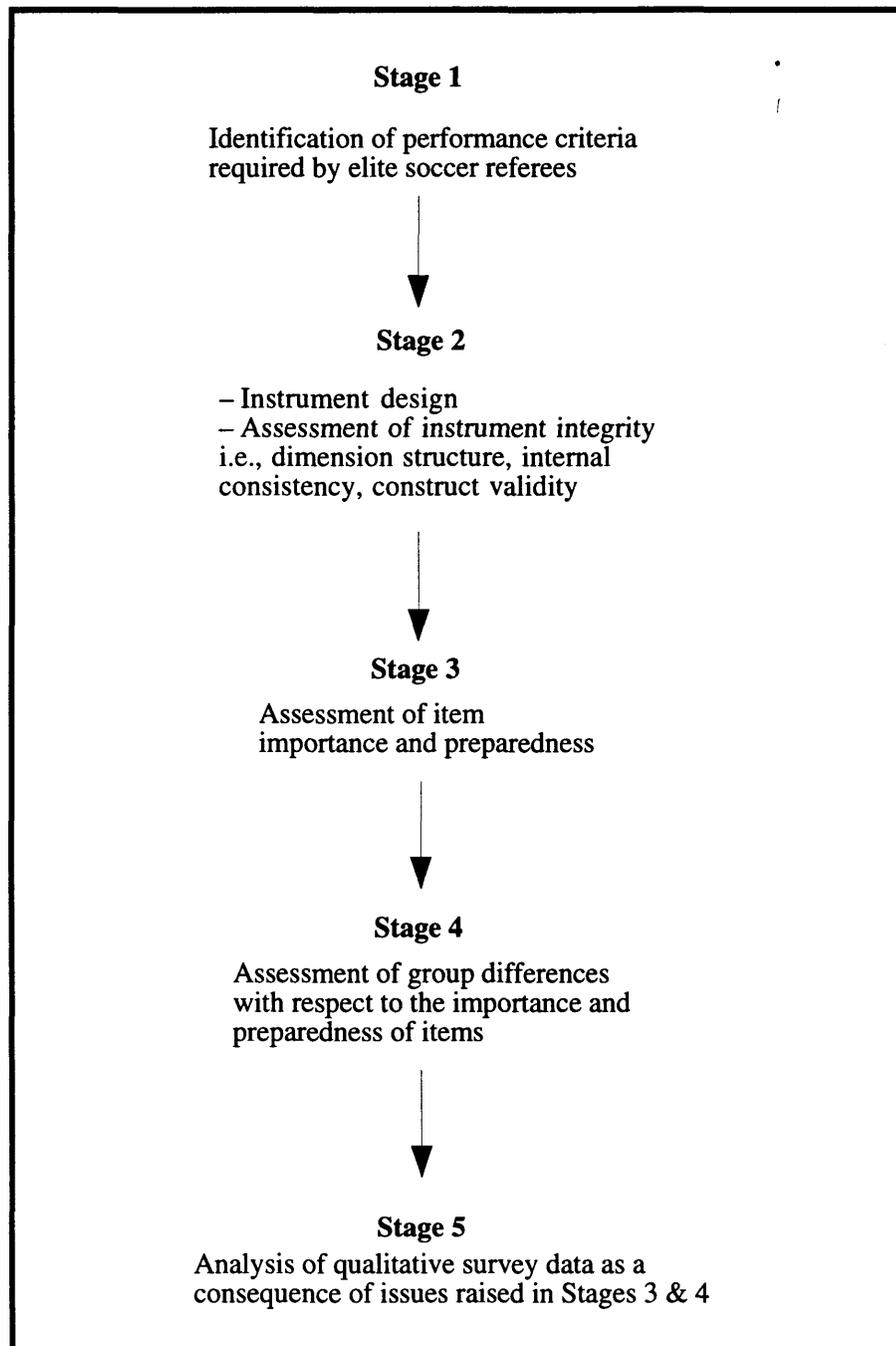
The study comprised five stages, with each stage consequential to the next. This design is summarised diagrammatically in Figure 4.1. In the discussion below, all stages are examined, with particular attention focused on the control of limitations and constraints identified previously.

Stage 1 encompassed the preliminary study, described previously (see Chapter 3). This stage identified the performance dimensions and competencies which were seen to be essential for elite soccer refereeing. It is these performance dimensions and competencies that formed the basis of subsequent instrument development and data analysis.

The second stage was designed to address a number of methodological issues, previously raised in Chapter 2, pertaining to the measurement characteristics of hybrid BARS instruments. The success, or otherwise, of confirming the essential nature of refereeing competencies, and their subsequent classification into performance dimensions, was central to the direction of further analysis (for specific details, see *Data Analysis Plan* later in this chapter).

The next three stages comprised a series of smaller investigations. In Stage 3, the performance criteria were assessed according to their perceived *importance*, and *preparedness* (this latter perspective is analogous to an overall assessment of how proficient elite soccer referees are in demonstrating essential performance criteria). Stage 4 assessed perceptions of players, coaches, referees, assistant referees and referee inspectors, with respect to the perspectives of *importance* and *preparedness*, to determine if significant differences of opinion exist between these groups. In Stage 5, issues arising from Stage 4 were explored through the analysis of qualitative survey data.

Significant to the last three stages was the identification of people who were qualified, either through their professional sporting experience, or status that they assume within the soccer hierarchy, to provide opinion on performance criteria across perspectives (i.e., the external validation process described in Chapter 2, see Anshel, 1995). Previous hybrid BARS studies have obtained opinion in similar ways, but have only included people who were



**Figure 4.1** Design sequence of the research study

involved specifically in the role under investigation (e.g., referees validating refereeing competencies, see Anshel, 1995; Anshel & Webb, 1991). However, such approaches limited the diversity of opinion which can be obtained on the role under investigation. This, in turn, constrains the validity of findings. In doing so, this presents data triangulation concerns (such issues are discussed in greater detail later in this chapter, see *Evaluation of Research Design*). Moreover, previous studies examined the importance of performance dimensions only, thus negating a full assessment of occupational competencies, i.e, how competencies are performed.

Consequently, the present study adopted a broader approach to this process, and included the opinions of people who, in undertaking their soccer-specific roles, are impacted by the referees ability to perform and execute refereeing skills proficiently. These people encompassed those who work directly with the referee in administering the rules of the game (assistant referees), are effected by refereeing decisions (players and coaches), or sit in judgement on referee performance (referee inspectors). This approach ensured that the most informed and diverse viewpoints were obtained about the importance and proficiency of referee performance criteria (a more detailed rationale concerning this approach is given later in this section).

Another important feature of the research design is the combination of qualitative and quantitative techniques. Stage 1 is characteristic of qualitative design, and the various qualitative aspects of this stage were reviewed in detail in Chapter 3. Stages 2, 3 and 4 moved away from qualitative methodology and embraced quantitative approaches. This research paradigm encompasses a number of techniques which enabled the investigator to quantify results (see Borg & Gall, 1989). Specifically, these stages assumed a descriptive predilection. With respect to this focus, descriptive research “systematically [describes] the facts and characteristics of a given population or area of interest, factually and accurately” (Isaac & Michael, 1981, p. 46), and is concerned primarily with finding out ‘what is’ (Borg & Gall, 1989). In the present study, such issues are implicit to the research questions and hypotheses. The final stage, Stage 5, reverted to the qualitative paradigm. Specifically, data collected in this phase is textual and non-numerical in nature. This type of data, and subsequent analysis, helped to provide answers to any ‘why’ questions which arise from earlier stages on the investigation.

It is pertinent to acknowledge here that the two paradigms were not viewed as mutually exclusive for the present study. On the contrary, their combined methodological strengths were seen as complimentary, thus enabling a thorough investigation of the research themes articulated in Chapter 2.

**Procedures**

Fundamental to the progress of the present study were four procedures. These procedures, built into the research design, were fundamental to the steady progress of the investigation. An overview of all procedural details, including the inter-relationship of design procedures to specific design stages (as listed in Figure 4.1) are specified in Table 4.1.

**Table 4.1 Chronology of the investigation procedures**

Procedural Step	Time	Activity	Research Design Stage
Preliminary Planning	2 months	<ul style="list-style-type: none"> <li>- Gain research ethics approval</li> <li>- Obtain support from Soccer Australia</li> </ul>	1
Development of refereeing competencies and indicators	3 months	<ul style="list-style-type: none"> <li>- Identification of expert panel members.</li> <li>- Meeting to determine refereeing competencies and indicators</li> </ul>	1
Development of instrument	2 months	<ul style="list-style-type: none"> <li>- Identification of sample</li> <li>- Identification of instrument dimensions (improvement, preparedness and improvement priority)</li> <li>- Use of Likert Scale format</li> <li>- Instrument pilot</li> </ul>	2
Collection of data:			2
a. - referees - coaches	One day	- One day meeting in Sydney used to collect data from referees and coaches	
b. - players	2 months	- Identification of research assistants	
c. - assistant referees - referee inspectors	2 months	- Postal questionnaires	

Initially, ethics approval from the University of New England’s Human Research Ethics Committee was required (see Appendix 6). Similarly, it was important to gain the support of Soccer Australia. As previously noted, Soccer Australia is responsible for the administration of the Ericsson Cup competition, and the enlistment of their support was both necessary and desirable. More specifically, as the study was concerned primarily with the competencies of elite referees, endorsement was needed from the National Referees Director at Soccer

Australia. Subsequent to verbal agreement from this person (pers. comm. Gary Power, 26 April 1996) to support this study, formal notification was obtained from Soccer Australia affirming official support for the study in any matters relating to the conduct of the investigation.

After gaining the required approval and support, it was essential to generate refereeing performance criteria prior to the new Ericsson Cup season. This task represented the second procedural step. It was essential for this step to utilise the 'out-of-season' period to maximise the range of people available to comprise expert panel membership. Had such timing not been observed, numerous logistical problems would have resulted, with many experts being unavailable due to playing or training commitments.

The third step in the research procedures incorporated instrument development. Although all detailed aspects of this step are provided in the following section, the instrument's development is worthy of comment in the context of the procedural time-line. Specifically, instrument piloting procedures presented greater timing flexibility than the performance criteria identification process. This was because the pilot sample was not involved directly in playing, coaching, refereeing or inspecting during the Ericsson Cup season. Thus, an 'out-of-season' focus was not as important for the pilot. Nevertheless, the piloting procedures still required completion by mid-season to facilitate data collection in the latter half of the season.

Collection of data was central to the fourth and final procedure. A variety of methods and strategies were adopted (see Table 4.1) due to a number of inter-connected reasons. In particular, the wide geographical disbursement of subjects, coupled with the previously identified indifference of players to formalised 'administrivia' necessitated collection strategies which varied among stakeholders. Specific details of each data collection technique used are detailed below:

#### Referees and Coaches

Collection of data from referees and coaches occurred over a one-day period, which coincided with a meeting between Ericsson Cup coaches and referees. The prime purpose of this meeting, convened by Soccer Australia, was to discuss major rule changes and rule emphasis for Ericsson Cup matches. However, in light of Soccer Australia's support for this investigation, officials from Soccer Australia invited the investigator to the meeting for the expressed purpose of data collection.

The gathering presented a unique opportunity to obtain data from particular sub-groups, which, under normal circumstances, would have been dispersed throughout Australia. All Ericsson Cup referees attended this meeting, as did 12 head-coaches and two assistant-

coaches. Consequently, at this meeting, the investigator was able to access a 'captured audience' for data collection.

### Players

In discussion previously (see *Planning for Research*), it was indicated that data from players were collected within the context of club training venues and times. To facilitate data collection, research assistants were identified in each city which had Ericsson Cup clubs. The research assistant's role was to collect player data at all Ericsson Cup clubs, and return the data to the investigator.

This process, however, was not without some problems. Despite the study being supported by Soccer Australia and the Australian Soccer Player's Association, not all clubs were willing to make players accessible to appointed research assistants. To overcome this barrier, people who had connections with, and acceptance by, such clubs, i.e., 'club-insiders,' were identified and recruited to act as research assistants. Although such an arrangement was not entirely desirable from a research perspective, it did meet with acceptance by those clubs who would not accept research assistants appointed by the investigator.

All research assistants (including those identified as club-insiders) were contacted by telephone and explained the purpose of the research. They were then sent information that detailed the data collection process (see Appendix 7). Follow-up telephone calls were made to ensure the data collection procedures were understood fully.

To avoid last minute problems with the collection of data, club officials were informed of the identity of the person who would collect data from the players. Moreover, each club *coach* was contacted by the investigator to inform them of the impending data collection. This contact was seen to enhance access at training venues, as the investigator was already known to all coaches (coaches had been through the data collection process previously, and were supportive of the study). All coaches indicated that access to players would not be obstructed.

Following this initial contact with clubs and coaches by the investigator, research assistants liaised with respective clubs to organise times which were mutually beneficial for data collection. All data were collected from each club before the end of the season to ensure the widest possible sample size.

### Assistant Referees and Referee Inspectors

Data from referee assistants and inspectors were collected via posting. Although the problems associated with this form of data collection were well documented (i.e., low return

rates, see Borg & Gall, 1989; Cohen & Manion, 1994; Drew, Hardman, & Weaver-Hart, 1996; Krathwohl, 1993), the disparate locations in which these people resided, coupled with a lack of regular and formalised gatherings of these groups (as opposed to the training requirements of players), suggested that postal data collection was the most viable and appropriate data collection method.

Postal addresses for both sample groups were obtained from Soccer Australia, and the survey instrument was sent to all assistant referees and inspectors. Other documentation accompanied the questionnaire. These were a 'Plain Language Statement' as stipulated by the UNE Ethics Committee (see Appendix 8), and a letter of transmittal (based on suggestions by Borg & Gall, 1989; Drew et al., 1996) explaining the purpose of the research and procedures for returning the data (see Appendix 9). After a three-week period, follow-up letters were sent out to encourage the return of surveys from subjects who had not already done so (see Appendix 10).

### **Summary**

In summary, the research design adopted for the present investigation embraced both qualitative and quantitative research paradigms. This approach was not only desirable from a research perspective, but fundamental to the exploration of research themes outlined in Chapter 2. Similarly, the diversity in research approaches is reflected in the variety of procedures adopted in the design. This was due primarily to the seasonal nature of elite soccer, and the need to access subjects who live throughout Australia.

## **SAMPLE SIZE AND INSTRUMENTATION**

This section provides discussion relating specifically to the sample size used in the present study, and aspects of instrument design and development.

### **Sample Size**

Sample sizes for each stakeholder group described previously is provided in Table 4.2. Figures from this table show sample sizes, with respect to each sub-group, are not equal. However, this is not unexpected given the nature of the groups being investigated. To clarify this point further, for each Ericsson Cup match there is a maximum of 28 players (11 players per team plus a maximum of three substitutes). This compares with one referee, two assistant referees, one referee's inspector, and a possible maximum of four coaches (two per team per game). Therefore, within the context of a soccer 'community' disparity in numbers across groups is typical.

More specifically, the vagaries of player selection did not allow a precise sample size to be determined. This occurred for two reasons. Firstly, a player excluded from the sample

**Table 4.2 Sample size by soccer-specific role and state of residence**

	Players <sup>1</sup>	Coaches	Referees	Assistant Referees	Referee Inspectors	Total
Queensland	14	2	2	5	4	27
NSW <sup>2</sup>	56	7	5	11	10	89
ACT	14	1	0	2	4	21
Victoria	56	7	2	7	5	77
South Australia	28	4	2	6	3	43
Western Australia	14	2	2	5	4	27
<b>Total</b>	<b>182</b>	<b>23</b>	<b>13</b>	<b>36</b>	<b>30</b>	<b>284</b>

during the planning phase of the investigation may, in fact, have played in the Ericsson Cup by the time data collection had begun. Secondly, player registrations obtained from Soccer Australia did not discriminate Ericsson Cup and Youth team registrations. However, it is estimated that Ericsson Cup clubs would have approximately 16 players under contract who would be expected to play in the Ericsson Cup during the season. However, the collection of data from playing squads of this size is optimistic given the practical realities of player injury, illness, and availability. Therefore, it is conservatively estimated that 14 players per club will comprise the potential player sample.

Similarly, the number of individuals from each sub-sample is different across States. Again, this is not unexpected due to the higher number of Ericsson Cup clubs in NSW and Victoria (thus explaining comparatively high player and coach numbers in these States). The variation in numbers of available referees, assistant referees and referee inspectors from smaller States add further to the inter-state differences.

### **Instrument Design and Development**

Discussion in this sub-section examines procedures that were undertaken for designing the research instrument, i.e., a questionnaire. Three aspects of questionnaire design are reviewed, and include design overview, questionnaire piloting, and questionnaire validity and reliability.

#### **Overview of Questionnaire Design**

A draft instrument was developed in accordance with questionnaire survey guidelines provided by Borg and Gall (1989). These included:

---

<sup>1</sup> Estimate of 14 players per club.

<sup>2</sup> Inclusive of Northern NSW, who are ordinarily considered a separate soccer 'state' by Soccer Australia.

- organising the layout of questions so that the instrument is as easy to complete as possible;
- numbering the questionnaire items and pages;
- including brief, clear instructions;
- using examples with items that might be confusing or difficult to understand;
- organising the questionnaire in some form of logical sequence;
- beginning with non-threatening items; and,
- providing enough information in the questionnaire so that items are meaningful to the respondent.

The draft questionnaire was divided into three sections. The first section included background information about each respondent. This allowed the respondent to begin the questionnaire with questions that were factual, non-threatening and simple to complete (as recommended by Borg & Gall, 1989). Moreover, the inclusion of such items allowed statistical comparisons needed to answer the stated research questions and hypotheses. As such, the first section contained forced-choice items which related to each respondent's age, state of residence, soccer status (i.e., their current role in the Ericsson Cup), occupational status, experience in their current Ericsson Cup role, education level, coaching and/or refereeing qualifications, and playing experience. Provision was made for respondents to include any other information about their soccer history which they saw relevant to this study.

The first section also requested that respondents supply their name. This was included to enable the investigator to contact respondents if clarification of any responses was required. However, the fulfilling of this request was optional, and respondents were free to ignore this request if they wished.

The second section of the questionnaire was designed to assess the *importance*, *preparedness* and *improvement-priority* of the refereeing performance dimensions and competencies identified in the preliminary study. The assessment of each item, with respect to these three perspectives, was achieved through the implementation of a rating format that incorporated a Likert scale configuration.

With reference to this configuration, the choice of Likert scales (also referred to as summated scales, see Bernardin, 1977; Bernardin, LaShells, Smith, & Alvares, 1976; Borman, 1979; Campbell et al., 1973) for assessing each item is contentious. There has been considerable discussion in BARS literature concerning the use of this format. One early empirical study

which compared various types of rating formats did not endorse Likert formats (Campbell et al., 1973). In that study, Campbell et al. (1973) compared traditional BARS formats (see Figure 2.1) with summated formats, and concluded the traditional BARS format was preferable in terms of method variance, leniency error, and halo error. However, subsequent research, which specifically compared format effects, concluded that BARS formats were not superior to Likert formats when assessed against a range of psychometric criteria (Bernardin, 1977; Bernardin et al., 1976; Borman, 1979). Moreover, Likert scales were claimed to improve the psychometric qualities of BARS instruments (Borman, 1979).

It is worth noting that three of the previously mentioned studies (Bernardin, 1977; Bernardin et al., 1976; Borman, 1979) generated items using BARS procedures. This may account, to a certain extent, for the lack of variability between the two formats. As deduced by Bernardin (1977), rigorously developed summated scales perform as well as rigorously developed BARS formats.

Although debate about the effectiveness of various BARS formats was inconclusive, the use of Likert scales *per se* was not without support. In particular, Likert scales were viewed as a useful for three reasons, namely: a method in measuring attitudes, beliefs and opinion (Borg & Gall, 1989; DeVellis, 1991); collecting data that is useable and analysable (Cohen & Manion, 1994); and, enhancing the internal consistency of an instrument (Oppenheim, 1992).

When measuring opinion or perception, the normal practise for Likert scales is to have a set number of response options per item. Usually, five options are provided, although four to seven options have also been used. Notwithstanding this variation, response options are designed to cover a continuum of opinion. For example, a four-option response continuum measuring agreement may include: 'strongly agree' – 'agree' – 'disagree' – 'strongly disagree.' For the present study, a five response option format, (as recommended by Landy & Farr, 1980) was adopted. Therefore, options for the *importance* and *preparedness* perspectives were defined on a continuum from 'very' to 'not' (i.e., 'very important' to 'not important,' and 'very prepared' to 'not prepared,' respectively), and the *improvement-priority* perspective was measured on a 'high' to 'low' continuum.

All response options were allocated a numerical value from one-to-five. However, a consistent numbering format – with respect to the direction of numerical values – was not provided in the BARS-related literature. Some studies have placed their most positive response option, e.g., 'definitely', or 'strongly agree,' with the smallest numerical value (for example, see Anshel et al., 1987; Anshel & Webb, 1991; Bernardin et al., 1976; Borman, 1979), while others have placed their positive response option with the highest numerical value (for example, see Anshel, 1995; Landy, 1985). As no standardised approach was

found in the literature, a decision was taken for the present study to place the most positive response option with the lowest numerical value, e.g., 'very important' was assigned a value of one, and 'not important' was assigned a value of five (for example, see Anshel & Webb, 1991). Although such numerical ordering is of little philosophical or conceptual importance, care was taken to ensure that ordering was consistent across perspectives.

The third and final section of the questionnaire contained one open-ended question. This provided an opportunity for respondents to add their own opinions and feelings about the issues under investigation. Specifically, respondents were encouraged to provide their impressions about the current standard of refereeing in the Ericsson Cup. This allowed supplementary qualitative data to be collected that complimented the quantitative data from Section 2 of the questionnaire.

Three other general design features are worthy of discussion. Firstly, to simplify the presentation of the instrument, all sections began on a new page. Secondly, each performance dimension and related competency, within Section 2, were grouped on a separate page. Lastly, the instrument was printed in a clear typeface with 10 point font to provide visual clarity to the questionnaire. These three features, consistent with the recommendations from Borg and Gall (1989), were seen to enhance the design and functionality of the instrument.

Overall, the design of the instrument was predicated on principles which served to improve the utility of the questionnaire. Although each design feature *per se* is not significant in this regard, the accumulated effect of all features fosters acceptance of the instrument, thus increasing the likelihood of the questionnaire being completed by the proposed sample.

### **Instrument Pilot**

Following the initial design and construction of a draft instrument, the instrument was piloted to assess a number of factors, including: the clarity of instructions; the clarity of items; and, the time and effort required on the part of the respondents in completing the questionnaire (Drew et al., 1996). Implicit in this assessment is the use of subjects who are similar in knowledge and experience to the research sample (Drew et al., 1996). Consequently, a pilot sample was drawn from people with previous soccer experience in refereeing, playing, coaching and inspecting at the national league level. Due to the highly specialised focus of this investigation, the pilot sample was relatively small ( $n=6$ ). It comprised two players, one coach, two referees and one referee inspector.

A two-hour pilot meeting was scheduled at the head-quarters of the NSW Soccer Federation. The procedure for the meeting followed piloting suggestions offered by Borg and Gall

(1989, p. 435). As such, respondents were asked to provide feedback relating to their understanding of the following issues:

- the meaning of each item. If ambiguity in meaning arose between any members of the pilot group, the item was reworded until the meaning of the item reached consensus. Care was taken not to alter the underlying meaning of any item;
- the intent and meaning of instrument instructions; and,
- their understanding of the perspectives being measured.

As a result of the meeting, a number of changes were suggested. These related to the amalgamation of some items, plus the rewording of others. This resulted in the original 38 competencies being reduced to 37. Additionally, there was one minor alteration in the instructions provided for referees. This was specifically related to the *improvement priority* perspective, which asked referees to rate the priority of each performance dimension and competency in respect to improving *their* individual refereeing. Other respondents, such as players, coaches, and referee inspectors, were given the more general instruction of “what *priority* would you place on each performance dimension and competency to *improve the quality* of refereeing in the Ericsson Cup ... ” These alterations, and the format of the final instrument, is provided in Appendix 11.

### **Instrument Validity and Reliability**

Chapter Two discussed, to some extent, the validity and reliability of previous instruments developed using hybrid BARS methodology (see Anshel, 1995). As indicated, findings were inconclusive. Nevertheless, it is important to assess the parameters of instrument validity and reliability for the questionnaire developed for the present study.

#### **Instrument Validity**

Instrument validity was examined from three perspectives, namely, *face validity*, *content validity*, and *construct validity*. Two other forms of validity associated normally with establishing the credibility of an instrument’s validity include *concurrent validity* and *predictive validity*. Although their importance is acknowledged, their relevance to this investigation is negligible, and are not discussed here.

With respect to the three forms of validity relevant to the present study, perhaps the most difficult to quantify is *face validity*. This form of validity is determined by a subjective judgement that the instrument “looks as though it would be valid” (Krathwohl, 1993, p. 203). To defend the face validity of the instrument used for this investigation, it is argued that all the items (except those in section one) related to the role of soccer referees. This is

reinforced through the procedural steps of the BARS process. Therefore, at least superficially, it is reasonable to assume the instrument is measuring what it was designed to measure, i.e., attitudes pertaining to the performance criteria of elite soccer referees.

Despite the apparent weaknesses associated with face validity, its impact and contribution to instrument integrity should not be ignored. As pointed out by Krathwohl (1993), people who are not knowledgeable about measurement judge an instrument by how it *looks*, rather than its content. For this reason alone, establishment of face validity is fundamental to instrument design. Moreover, when used in association with other forms of validity, face validity reinforces the overall acceptance of an instrument (Isaac & Michael, 1981).

The second form of validity requiring examination, *content validity*, is basically a representation problem (Krathwohl, 1993), and is concerned with the degree instrument items represent a specific content domain (DeVellis, 1991). Although somewhat similar to face validity, content validity is determined more systematically. This involves defining the specific content universe to be sampled, specifying objectives, and describing how the content universe is sampled to develop instrument items (Borg & Gall, 1989; DeVellis, 1991). The hybrid BARS procedures used for this research – as described in Chapters 2 and 3 – aided in the assignment of content validity to the instrument developed for this investigation (see Anshel, 1995). Specifically, the use of diversified expert panels, who were under instruction to develop the performance domains and competencies related to elite soccer referees only, ensured the content area under investigation was focused appropriately.

The associated advantages of using occupational experts for enhancing content validity is strongly supported by DeVellis (1991). He saw experts providing three unique advantages with respect to establishing content validity, namely, the suitability of experts to:

1. determine the relevance of each item;
2. evaluate an items clarity and conciseness; and,
3. identify items that comprise the entire content domain.

The BARS process complimented these benefits further through the inclusion of all items developed by each respective expert panel. This negated the omission of items (that commonly occurred in the development of traditional BARS scales) which, in turn, contributes to the rigour of questionnaire content validity.

A third area of validity, *construct validity*, examines the extent to which an instrument measures a hypothetical construct (Crowl, 1993). On occasions, traditional BARS studies have used factor analysis techniques successfully to support the construct validity of BARS instruments (Kafry et al., 1979; Kavanagh & Duffy, 1978; Keaveny & McGann, 1975;

Stoskopf et al., 1992). In these studies, the competencies were analysed to see if they formed interpretable factorial solutions. If such a solution was found, the construct validity of the instrument was claimed to be verified, i.e., factors were seen to be analogous with performance dimensions. This study sort to establish construct validly in the same way, although the determination of this form of validity is dependent on data collection and analysis. Therefore, results pertaining to construct validity are addressed in Chapter 5.

Notwithstanding the establishment of construct validity, the instrument designed for the present study exhibited sound credentials with respect to relevant validity criteria. This is consistent with other BARS instruments, in both traditional and hybrid forms. Particularly, the inherent design procedures of BARS address validity issues, and are viewed as one of the methodology's prime strengths with regard to instrument development.

#### Instrument Reliability

Generically, reliability is a statistical concept "based on the association between two sets of scores representing the measurement obtained from the instrument when it is used with a group of individuals" (Wiersma, 1995, p. 309). Ordinarily expressed in the form of a correlation co-efficient, reliability can be interpreted from a number of perspectives, including *internal consistency*, *equivalence*, and *stability*. The latter two forms of reliability are not applicable for this investigation, as they require either two or more parallel forms of the instrument (equivalence), or 're-testing' the same sample (stability).

However, reliability measures of *internal consistency* are pertinent. This measure assesses the relationship of instrument items to each other, and an underlying scale continuum (Oppenheim, 1992). Implicit in the determination of this measure is verifying the performance dimension structure (i.e., construct validity). If the structure is sustainable, internal consistency can be established within each dimension. Conversely, verification is required independently of the dimensions if the dimension structure is not sustained. To assess internal consistency via item-relationship, a Cronbach's Alpha co-efficient was computed, and is reported in Chapter 5.

A second aspect of internal consistency was related to responses across the perspectives of *importance*, *preparedness*, and *improvement-priority*. Conceptually, the difference between the importance of an item, and the item's preparedness, should represent the degree of required improvement, i.e.,

$$\text{Importance} - \text{Preparation} = \text{Improvement priority}$$

As such, scores were obtained which represented the difference between *importance* and *preparedness*, which, in turn were correlated with *improvement priority* scores. A high

correlation coefficient ( $\geq 0.70$ ) would indicate respondents were interpreting each of the perspectives as intended.

For this to be determined, a Pearson Product-Moment Correlation coefficient was used to calculate the extent of such a relationship. Pearson's correlation is widely used, and requires both variables to be expressed on an interval scale (Burns, 1997). However, in the present study the Likert scale format adopted for the questionnaire presented ordinal level data. Under normal circumstances, this would present a statistical concern. To address and overcome this problem, a statistical technique can be employed to 'convert' ordinal to interval data, i.e., Rasch item analysis (see *Data Analysis Plan* for specific details). It is this converted data which were used to calculate Pearson's correlation. Results for this measure of internal consistency are also presented in Chapter 5.

Notwithstanding the results of internal consistency calculations, the format of the instrument can also improve internal consistency. As indicated by Borman (1979):

One potential advantage [of Likert scales] is that making several ratings rather than one per dimension ... may increase the reliability of the dimension score. This increase in reliability may emerge for the same reasons that adding items to a test generally increases the test's reliability.

In relation to Borman's quote, the format of the final instrument encompassed 37 competencies, spread across six performance dimensions. This ensures that multiple measures of performance are obtained, and, in turn, this improves the reliability of the instrument over-and-above those levels that would be achieved if only performance dimensions had been measured.

### **Summary**

The instrument used in the principal study was predicated on refereeing performance requirements identified in the preliminary study. Primarily, the instrument was designed to assess each performance criteria against the perspectives of importance, preparedness, and improvement priority. The Likert scale format was adopted to measure the magnitude of these perspectives.

Although all validity and reliability parameters of the instrument could not be assessed here, judgements were possible concerning face validity, content validity, and internal consistency. With respect to these three measures, the research instrument demonstrated acceptable properties. Judgements relating to the performance dimension structure, plus additional measures of internal consistency, are reported in Chapter 5.

The sample chosen to complete the survey instrument is large ( $n=284$ ), and comes from across Australia. Although the bulk of the sample are players, for reasons mentioned previously this was not unexpected. The disparate size of sub-groups in the sample is unavoidable in research of this nature, and is an issue addressed in greater detail later in this chapter (see *Data Analysis Plan*).

## EVALUATION OF RESEARCH DESIGN

This section considers the design with respect to the possible impact of extraneous variables. Implicit in this discussion is the delineation of techniques employed to control factors that lie outside the research design. Specifically, these relate to a number of idiosyncratic features associated with the sample, and include: the age of respondents; general soccer experience; specific Ericsson Cup experience; attitudes towards referees and refereeing; understanding of English (the multicultural nature of soccer suggests that English may be a second language for some respondents); educational status; and, the indifference of some respondents to completing tasks which are not of a practical nature.

In view of these factors, a number of measures were suggested which acted to limit the erroneous effect of extraneous variables. Various philosophical stances underpin these actions. However, some perspective is placed on this approach by Patton (1990, pp. 38-39), who stated:

Rather than believing that one must choose to align with one paradigm or another, I advocate a paradigm of choice. A paradigm of choices rejects methodological orthodoxy in favour of methodological appropriateness as the primary criterion for judging methodological quality. The issue then becomes ... whether one has made sensible methods decisions given the purpose of the inquiry, the questions being investigated, and the resources available.

In light of these comments by Patton, the appropriateness of the research design for the present study is considered. Discussion takes into account the extraneous variables impacting on the study, and provides a perspective on the design from two seemingly dichotomous research paradigms, i.e., traditional (experimental and quasi-experimental) quantitative designs, and qualitative (naturalist and constructivist) designs. Although these research paradigms are not mutually exclusive, they have been seen historically as contrary, and, as such, provide diverse criteria for assessing the integrity of the chosen research design.

### Quantitative Paradigms

The customary benchmark for establishing the appropriateness of research design is *validity*. It is central to the effectiveness of any investigation, and is assessed against a number of predetermined criteria (although it should be acknowledged that not all criteria are applicable

to every design). Campbell and Stanley (1963), in what many sources considered the foundation paper on this topic (see Borg & Gall, 1989; Cohen & Manion, 1994; Isaac & Michael, 1981; Miller, 1991), cited 12 variables which can impact on the validity of a research design. The variables were further classified into the categories of *internal* and *external* validity, comprising eight and four variables, respectively.

### **Internal Validity**

Internal validity can be defined as the “degree to which the research findings can be distorted by extraneous variables” (Borg & Gall, 1989, p. 405), and was seen by Campbell and Stanley (1963, p. 175) as the “*sine quo non*” of research design. However, a number of threats to internal validity are not applicable to the design of this research. These include: history; testing; maturation; statistical regression; and, experimental mortality. However, two threats, referred to as instrumentation and sample selection, are applicable, and their potential impact is discussed below.

*Instrumentation:* Data gathering instruments which are unreliable may introduce errors into the data. In order to minimise such a threat in the present study, instrument development techniques which were seen to be reliable and acceptable were implemented. Briefly, these techniques required the development of specific performance criteria across the entire occupational domain of elite refereeing responsibilities, thus ensuring complete coverage of the refereeing role. Moreover, the design of the investigation attested that instrument format was thoroughly piloted for content and clarity.

*Sample Selection:* The selection of subjects can bias results, i.e., bias resulting in the differential selection of respondents for a comparison group. Although Campbell and Stanley (1963) viewed this variable from an experimental perspective, the use of specific sub-groups within the global sample employed for the present study made issues of sample bias relevant.

In addressing this issue for the present study, potential problems with selection bias were effectively limited by all-inclusive sampling, i.e., all potential subjects were included in the sample. Such inclusion was not only optional, but desirable. As indicated previously in this chapter, various subgroups of the sample were small, thus negating any need for cross-sectional sampling. Even the largest sub-group (i.e., players), was not reduced for data collection purposes. As a result, the focus on global sampling reduced markedly the chances of sampling bias occurring.

## External Validity

External validity has been specifically defined as the extent to which observations and measurements can be compared accurately with other groups (Kincheloe & McLaren, 1994), i.e., can results be applied in other contexts or to other relevant populations. Under the umbrella of traditional experimental research paradigms, the threats to external validity identified by Campbell and Stanley (1963) were later classified into two main components of external validity, namely, *population validity* and *ecological validity* (Bracht & Glass, 1968).

*Population validity:* This validity criteria is concerned with the degree to which results of an investigation can be generalised from the experimental sample to a larger group of subjects (Borg & Gall, 1989). In this regard, two forms of population validity emerged from the work of Bracht and Glass (1968). The first related to the extent one can generalise from the experimental sample to a defined population. Under conventional experimental formats, it is standard practice for the design to randomise the sample, and apply results from the sample to a specific population. This is despite the application of findings in this manner being considered risky (Borg & Gall, 1989). However, in the present study, this threat was overcome by engaging global samples from each subgroup. This facilitates the development of firm conclusions, as the sample population and the target population are the same.

The second aspect of population validity is the extent to which the personal variables of sample members interact with treatment effects (Borg & Gall, 1989). As this research design does not incorporate treatments, this category of population validity is not applicable to the present study.

*Ecological validity:* This form of validity encompasses “the extent to which results of an experiment [investigation] can be generalized from the set of environmental conditions created by the researcher to other environmental conditions” (Borg & Gall, 1989, p. 650). With respect to the present study, this application of validity is somewhat problematic because factors which contribute to this threat are, once again, more related to traditional experimental designs. Nevertheless, the issues raised through ecological validity are worthy of discussion. In the context of this study, the environmental conditions may be designated as the level of soccer under investigation, i.e., elite Ericsson Cup standard. While it would be inappropriate to compare the findings of this study to other levels of soccer (e.g., that played at amateur district), results may be applicable to other semi-professional or professional soccer leagues. Although comparisons would require careful consideration, they would be more sustainable if the playing and officiating standards were similar to the Ericsson Cup. Likewise, any valid comparison would also need to be considered in the context of the associated cultural and social mores of each comparative setting.

From the discussion thus far, it can be seen that the design for the present study is not entirely applicable to many of the traditional techniques which assess research design. This is particularly true for those methods tailored specifically to research designs with treatment and control/treatment samples. In light of this, it is appropriate to assess the integrity of the design against qualitative criteria. As substantial translations of validity concepts do exist (Drew et al., 1996), examination of validity issues from a qualitative perspective are not only possible, but desirable.

### Qualitative Paradigms

In recent years, alternative criteria to establishing the *bona fide's* of qualitative research design have gained greater acceptance. The alternative focus emanated from a construct Guba (1981) called 'trustworthiness,' and borrowed heavily from the ethnographic approach to research. In the establishment of trustworthiness, four fundamental questions require addressing to validate the research design (Guba, 1981). Specifically, these included:

1. How can one establish confidence in the 'truth' of the findings of a particular inquiry with respect to subjects and the context in which the inquiry was conducted?
2. Does the inquiry have applicability in other contexts and to other samples?
3. Can it be determined that the findings of the inquiry would be consistently repeated if the inquiry were to be replicated within a similar context and with similar subjects?
4. Can it be established that the findings of an inquiry are based solely on subject responses, and not the biases, motivations, perspectives, etc, of the investigator?

From within these questions, the ethnographic criteria of credibility, transferability, dependability, and confirmability were derived (Denzin & Lincoln, 1994). These criteria provided a detailed and comprehensive framework for assessing the integrity of particular research designs, such as the present investigation. The following discussion addresses these four criteria, and in doing so, draws heavily on the writings of Lincoln and Guba (1985). This source offered numerous techniques to address each criteria, although, as with the establishment of validity through traditional methods, some techniques are not applicable to the present study. Consequently, only techniques relevant to addressing threats to the design of this study are addressed in the following discussion.

### Credibility

Credibility is the traditional scientific parallel of internal validity (Guba, 1981). Inasmuch, a variety of methods are also used in its establishment. Firstly, it is suggested that the

investigator employ *prolonged engagement*, i.e., “soaking in the culture through his or her pores” (Lincoln & Guba, 1985, p. 302) with, and in, the research context. This allows the researcher to become *au fait* with the nature of the phenomenon under investigation. For the present study, the investigator met this criteria through extensive involvement as a former player and assistant coach at the Ericsson Cup level. Moreover, playing and coaching at this level aided in the acceptance and trust of the investigator by those under investigation. As noted by Lincoln and Guba (1985, p. 302), “It seems likely that unless the inquirer began as an accepted member of the group or agency being studied, distortions can never be overcome ... ”

Secondly, the technique of *triangulation* (now an integral part of qualitative design, see Borg & Gall, 1989; Cohen & Manion, 1994; Hakim, 1987; Marshall & Rossman, 1989) can enrich the credibility of research findings. Triangulation encompasses multiple measures being taken to “ensure that the variance reflected is that of the trait or treatment and not that associated with the measures” (Huberman & Miles, 1994, p. 438). Embedded in the concept of triangulation are a number of associated techniques, with the most common being *data triangulation* (Krathwohl, 1993, p. 328). This technique involves the use of two or more sources to establish accuracy (Krathwohl, 1993, p. 328). This practice is addressed by the present study through obtaining data from five sources, namely: players, coaches, referees, assistant referees and referees inspectors. As such, a variety of data sources was able to provide a diverse range of perceptions on the subject under investigation.

*Peer debriefing* is a third technique which assists in the establishment of credibility. Debriefing requires the investigator to communicate his or her research thoughts to a disinterested peer (Lincoln & Guba, 1985). In doing so, this process serves a number of purposes, and includes:

1. keeping the investigator ‘honest’ through the probing of an experienced protagonist playing the devil’s advocate. This allows the investigator’s bias to be probed, meanings explored, and the basis for interpretations clarified;
2. providing an opportunity to test working hypotheses;
3. providing an opportunity to develop, and test, approaching steps in methodological design; and,
4. giving the investigator the opportunity to clear the mind of emotions and feelings that could be clouding good judgement.

Although there is no set formula for the debriefing process, it is recommended that certain criteria be observed with respect to the debriefer, including: being a peer of the investigator; knowledgeable about the substantive area of the investigation; understanding of the

methodological issues employed in the research design; and, especially pertinent to this study, should not be a doctoral supervisor of the investigator (Lincoln & Guba, 1985, p. 309). In light of these recommendations, the investigator for the present study sought the assistance of a senior academic in the field of physical education and sports studies. This person was also cognisant with the BARS methodology described in Chapters 2 and 3, and so proved an ideal person to undertake debriefing responsibilities.

### **Transferability**

According to Lincoln and Guba (1985), all that is necessary in classical traditional research paradigms to establish transferability (external validity) is ensure 'Sample A' results can be generalised to a similar population. However, from a qualitative perspective, the transferability of findings from one study to the next is problematic (Marshall & Rossman, 1989). This is particularly true if, as in the present study, the investigation is context specific. In meeting this challenge, it is contended that it is not the responsibility of the investigator to provide in *index* of transferability, but rather to provide a *data base* that makes transferability judgements possible (Lincoln & Guba, 1985, p. 316). To achieve this, the investigator needs to set out working hypotheses and "thick descriptions" (Lincoln & Guba, 1985, p. 316) of the time and context in which the study are to unfold. This study inherently addressed this issue as a doctoral thesis. The procedures, context, design, data analysis procedures, and findings are all provided in comprehensive detail. Subsequently, studies which succeed this investigation are provided with detailed descriptions of all phases of the investigation. This will aid future investigators to transfer design decisions into other research settings. Therefore, the application of methodology and findings to other settings is a judgement incumbent on the next investigator.

### **Dependability**

Dependability is concerned with the stability of data (Guba, 1981). Although a number of qualitative techniques are provided in the literature to establish dependability (Guba, 1981; Lincoln & Guba, 1985), the most pertinent of these to the present study is the *overlap method*. This method, whereby different techniques are used to generate data and derive results, was advocated to overcome inadequacies in individual techniques (Guba, 1981). Dependability was established for the present study through the use of expert panels to generate refereeing competencies and performance indicators, then by using a wider sample to verify the data were, in fact, essential and important. The results of this latter check are quantifiable, and are presented in Chapter 5.

### **Confirmability**

This criteria is related to the objectivity of data (Guba, 1981), i.e., did the investigators attitudes and prejudices influence findings? The present study addressed confirmability by the removal of the investigator from most forms of data collection. This was particularly so in the process used to identify elite refereeing performance criteria. As noted in Chapter 3, the generation of these data were accomplished through the interactions of soccer-related experts, who in turn were facilitated by panel-leaders. The investigator was not privy to this process, thus negating the possible 'contamination' of data through investigator bias.

### **Summary**

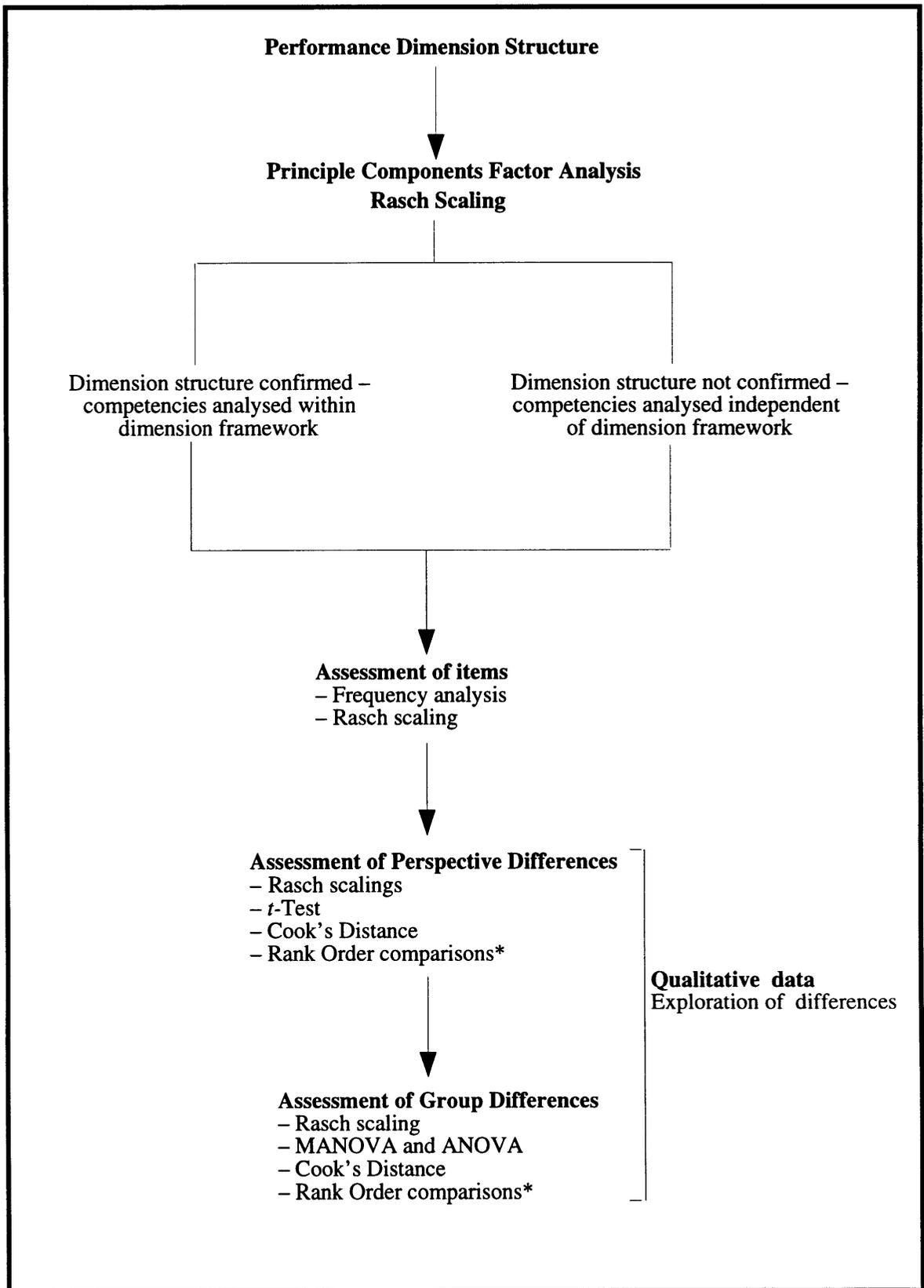
In conclusion, the previously cited quote from Patton (1990) at the beginning of this section (p. 104) showed that research designs can reflect a "different strokes for different folks" approach (Lincoln & Guba, 1985, p. 294). Accordingly, methods employed to establish design rigour are dependant on the nature of the investigation. The research design for the present study, employing design aspects from qualitative and quantitative paradigms, needed to address design threats from both perspectives. Implicit in such an examination is an acknowledgment of the relevant threats to the research design and the consequential selection of appropriate methods for negating threats, regardless of paradigm classification for the method.

## **DATA ANALYSIS PLAN**

Analysis of the data collected in the present study may take one of two possible paths. As shown in Figure 4.2, the direction of analysis is dependent on establishing the integrity of the competency structure derived from the BARS process. This was examined through two techniques, namely, factor analysis and Rasch scaling. If it is determined that the competency structure is not statistically sustainable, data were analysed independently of the competency structure.

### **Factor analysis**

The first technique used in assessing the validity of the competency structure is factor analysis (see Figure 4.2). Factor analysis procedures detail "underlying patterns or relationships for a large number of variables and to determine whether or not the information can be condensed or summarized in a smaller set of factors or components" (Hair, Anderson, Tatham, & Black, 1995, p. 365). It is an interdependence technique in which all variables are considered simultaneously, with each variable related to each other (Hair et al., 1995). For the purposes of this investigation, the competencies represented variables, from which



\* Used only if Cook’s Distance does not detect significant outliers.

Figure 4.2 Pathway for data analysis

underlying factors are sort. These factors are analogous to performance dimensions identified through the BARS procedure.

Specifically, the present study utilised the Principal Components form of factor analysis, with oblique OBLIMIN rotation. This procedure makes a number of assumptions, the most pertinent of which is the presumed interdependence of factors (Tabachnick & Fidell, 1996). In the context of this investigation, it is likely that the components which comprise the role of elite refereeing are not mutually exclusive and, therefore, any analysis should take account of this inter-relationship. Resulting factorial solutions were assessed using the unambiguous eigenvalue criteria of greater than or equal to one. Solutions were also assessed using two less objective measures, namely, the scree criteria, and, secondly, the application of the factor structures to appropriate practical settings. Detailed examination of these procedures is detailed in Chapter 5.

A second statistical method also used to assess the competency structure was Rasch scaling. Specific characteristics of this technique are discussed later in this section, as Rasch techniques were also employed for other analytical operations. However, with respect to the competency structure, it is important to note that Rasch analysis assumes homogeneity of data. Subsequently, if data are shown to fit the Rasch model, it can be deduced that the data are unidimensional and thus devoid of latent traits (analogous to performance dimensions).

### Frequency Analysis

Notwithstanding the construct results of factor and Rasch analysis, a number of techniques were employed to assess the importance and preparedness of items. Firstly, frequency analysis, reported in the form of cumulative percentages, has become the standard form of hybrid BARS data analysis (Anshel, 1995; Anshel et al., 1987; Anshel & Webb, 1991; Jessup, 1994). Although this form of analysis is somewhat simplistic, it does provide an indication of the level of agreement among respondents concerning the importance and preparedness of BARS items.

With respect to the *importance* of items, previous studies have used 90% agreement between respondents as a benchmark for the classification of a performance dimensions. Items which were rated either a 1 or 2 by 90% of respondents was considered 'must-have,' while those competencies which did not make this criteria, but where rated either 1, 2 or 3 by 90% of respondents, where considered 'should-have.' Any items which did not make either of these two criteria were considered to be 'unimportant' or 'non-essential' (Anshel, 1995; Anshel et al., 1987; Moore et al., 1997). For the present study, the following classification categories are used with respect to each perspective: importance (must-have, should-have and

unimportant); and, preparedness (very-well, well and poorly). In doing so, the 90% agreement rate was used as a benchmark for item acceptance.

As a result, the 90% benchmark provided a proxy validation for the selection BARS items. This was because items developed by the BARS panels were seen to be *essential* performance criteria required for successful refereeing. As such, agreement about the importance of items is a form of validation of the expert panel's perception about the inherent essential nature of each item.

The frequency analysis criteria applied in this study also allowed the detection of variability in group responses. As the present study is designed to seek a variety of opinions from different groups of subjects, the frequency analysis process can be used to explore variations in response patterns between these groups. Although empirical differences can not be established unequivocally through this technique, it does provide an indicator of possible group differences which, in turn, can be pursued using more sophisticated analytic techniques.

### **Rasch Scaling**

Detailed parametric analyses have not been applied previously to hybrid BARS data. Reasons for this are not apparent from the literature, although it could be hypothesised that the relatively homogeneous nature of cohorts from previous investigations precluded such analysis (see Anshel et al., 1987; Anshel & Webb, 1991; Moore & Webb, 1995). Moreover, data derived from Likert scales are ordinal in nature, and, therefore, the magnitude of the differences between adjacent points on the scale is unclear. Statistically, this would preclude further analysis using more powerful parametric techniques. However, Likert scale data have been analysed in the past by parametric techniques – specifically ANOVA – through using marginal totals as dependent variables (obtained by summing across rows and down columns). This was despite such techniques being of doubtful validity, particularly given the ordinal nature of data (Kerlinger, 1986).

Notwithstanding these concerns, data of this kind can be analysed using the Australian Council for Educational Research's QUEST software (Adams & Khoo, 1996). This analytic tool incorporates an implementation of the Rasch latent trait scaling model. Basically, the model converts ordinal scores into linear scores, thus facilitating the direct interpretation of data (Burton & Miller, 1998; Wright & Linacre, 1989).

Specifically, the model is based on both order and objectivity, the two fundamental principles of measurement (Hands, Larkin, & Sheridan, 1997; Snyder & Sheehan, 1992). The model is able to provide estimates of item difficulty (analogous to item importance, preparedness and improvement priority in the present study) and respondent ability (analogous to the

subject's perceptions) for polychotomously scored items (Wright & Masters, 1982). Importantly, the estimates of item difficulty and respondent ability are expressed on a logit scale and hence as an interval/ratio measure (Hands et al., 1997). Consequently, estimates can be utilised for subsequent empirical analysis techniques in assessing differences across perspectives and groups.

Conceptually, the determination of item difficulty is accounted for by whatever 'abilities' the persons in the sample happen to demonstrate. This frees the item difficulty estimates from these respondent 'abilities' (Snyder & Sheehan, 1992). Item difficulty is defined as the point on the scale where 50% of the sample responded 'correctly,' and, in a similar manner, the influence of item difficulty estimates are removed from estimates of person ability. Person ability scores are therefore a result of transforming raw scores to the same logit scale used to measure item difficulty (Snyder & Sheehan, 1992).

Although a number of Rasch model variants exist, the present study specifically employed the *Rating Scale* form of the Rasch model, which represents a constrained version of Rasch's Partial Credit model. The Rating Scale model is recommended for the analysis of attitudinal data (Andrich, 1982; Wright, 1998; Wright & Masters, 1982), and, in particular, where a set of items share the same rating scale structure (Wright, 1998). In this regard, a number of important assumptions need to be met. Firstly, the model assumes the same set of rating points is used with every item. Secondly, "the relative difficulties of the steps in each item should not vary from item to item" (Wright & Masters, 1982, p. 48). Lastly, the third assumption is closely aligned with the second, and requires that the magnitude between adjacent points on the Likert scale are not equal. Inasmuch, the Likert scale format meets this set of assumptions.

Algebraically, in the Rasch Rating Scale model, the probability of person  $n$  responding to category  $x$  to item  $i$  is given by:

$$\pi_{nix} = \frac{\exp \sum_{j=0}^x [\beta_n - (\delta_i + \tau_j)]}{\sum_{k=0}^m \exp \sum_{j=0}^k [\beta_n - (\delta_i + \tau_j)]}$$

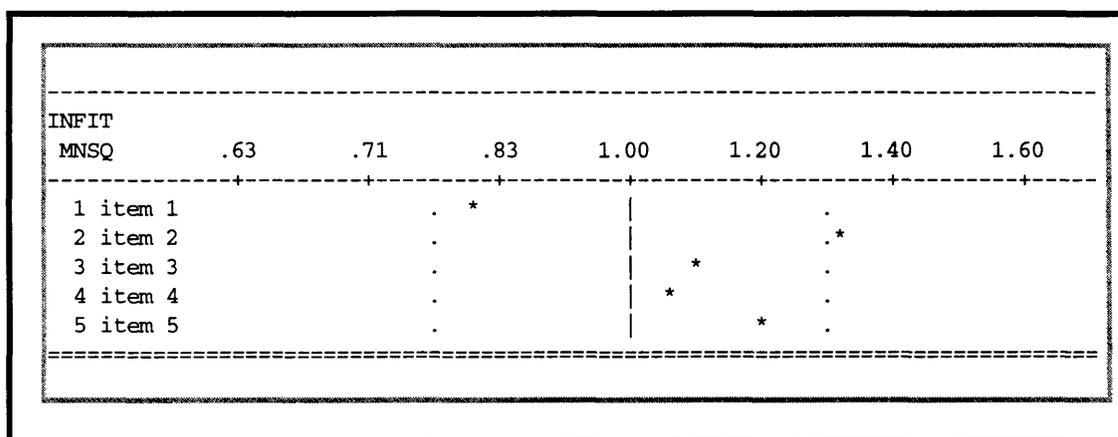
where  $\tau_0 \equiv 0$  so that  $\exp \sum_{j=0}^0 [\beta_n - (\delta_i + \tau_j)] = 1$ .

When this model is applied to the analysis of a rating scale, a position on the variable  $\beta_n$  is estimated for each person  $n$ , a scale value  $\delta_i$  is estimated for each item  $i$ , and  $m$  response 'thresholds'  $\tau_1, \tau_2, \dots, \tau_m$  are estimated for the  $m+1$  rating categories.

The suitability of data to the model is fundamental. If the data can not be managed by the model, then they can not be used to calibrate items or measure persons. In this regard, the

requirements for measuring are specified in the model, and can be analysed by three measures. Firstly, fit statistics (for both item estimates and person estimates), are provided as unweighted (outfit) and weighted (infit) residual based statistics. The QUEST software expresses each of these statistics as a mean-square and t-value. When data are compatible with the model, the expected values of the mean squares is approximately 1.0, and the expected values of the t-values is approximately zero (Adams & Khoo, 1996).

Secondly, the item fit map, produced by the QUEST software, shows the infit mean square for each item. This map includes two vertical dotted lines which enclose items with acceptable values. The software uses an arbitrary measure of 30% above and 30% below the expected item values for these vertical lines (Adams & Khoo, 1996). This allows the level of parameter fit in the model for each item to be assessed. In doing so, it provides a visual representation of the degree of homogeneity in the data. An illustration of an item fit map, using five items as an example, is provided in Figure 4.3.



**Figure 4.3** Example of Rasch item fit map

Thirdly, the issue of unidimensionality is given further empirical support by an *item consistency index*. The index measures the extent to which the items are homogeneous, and, in this regard, can be considered analogous to Cronbach's alpha (Adams & Khoo, 1996, p. 45, 93). It is this measure, coupled with the infit-mean square map illustrated in Figure 4.3, which is used in conjunction with factor analysis procedures in determining the validity of the BARS dimension structure (see Figure 4.2).

Lastly, the degree of success in defining a continuum depends on the extent to which the items and persons are separated (Wright & Masters, 1982). Items and persons must be sufficiently well separated in order to provide direction and meaning to a variable. The *reliability of estimate*, obtained through QUEST as a function of Rasch analysis, provides this detail. The estimate is the proportion of the observed variance that is considered true (Adams & Khoo, 1996), and provides a measure to assess the separation of the items and subjects. In practical terms, this represents the likelihood of an item's or a person's position

on the continuum remaining constant. Estimates above 0.7 are considered acceptable. Consequently, firm conclusions can be made about the relative position of each item and person (or group of people), a process which is not possible with simple frequency analysis procedures.

### **Parametric analysis – MANOVA, ANOVA, *t* Tests, and Cook's Distance**

As noted earlier, item and case estimates obtained from QUEST are given in the form of interval scores. As such, these scores can be used in parametric tests to determine significant differences between perspectives (*importance* and *preparedness*) and groups (referees, assistant referees, referee inspectors, players and coaches).

#### **Perspective Differences**

For assessing perspective differences, data from both perspectives are submitted to a single Rasch scale to obtain item estimates. This analysis is required initially to convert ordinal data (derived from the instrument Likert scales) into interval data. Each item, and corresponding item estimate, are sorted back into perspectives. The means and standard deviations of item estimates, from within each perspective, are then submitted to a paired sample *t* test using the SPSS statistical package. Given a significant difference, Cook's distance, a form of regression analysis (Stevens, 1996), is implemented to detect which individual items are most implicated in causing the significant difference. If Cook's distance fails to detect significant outliers, non-parametric rank-order differentials are used to indicate items most implicated in perspective differences.

#### **Group Differences**

The analysis of group differences is less straight forward, mainly due to the relatively small potential sample sizes of the referee ( $n=13$ ) and the coach ( $n=18$ ) sub-groups. Therefore, a decision was taken to collapse the five sub-groups into two groups. Consequently, the referees, assistant referees and the referee inspectors are formed into an 'officiating group,' while the players and coaches are classified into a 'competitive group.' This gives each group potential sample sizes of 79 and 205 respectively.

The grouping decision, essentially made on statistical grounds, has theoretical and practical foundations also. As outlined in Chapter 1, the motivations and game objectives of players and coaches are similar, and in this regard seem dichotomously opposed to the motivations of the officiating group. For this reason, it could reasonably be assumed that the opinions of players and coaches are somewhat homogeneous, as would the opinions of referees, assistant referees and referee's inspectors.

Case estimates, derived from Rasch analysis, provide the data for analysis. Group differences across perspectives are assessed initially through MANOVA, computed using the SUPERANOVA statistical package, to determine the existence of possible significant differences occurring. If a significant difference is found, step down one-way ANOVA can determine in which perspective group differences arise. Moreover, the Cook's distance technique (discussed previously with respect to uncovering 'outlier' items for perspective differences) can be employed to determine which items are most related to group differences. If Cook's distance fails to detect significant outliers, rank-order differentials are used to indicate items most implicated in group differences.

### **Qualitative analysis**

Data from the open-ended question were analysed using the NUD•IST software package. This software facilitates the indexing, searching and theorising of unstructured and non-numerical data (QSR, 1996). The software builds two inter-related databases. The first of these is the document database which stores all textual and non-textual data. Each subject's response is transcribed, and held as a separate document within this database. From this, the investigator is able to store and retrieve the text of any document, and search for words and phrases in the text of any document and automatically index these (QSR, 1996). Indexed information is stored in the second database. The index database allows the creation of categories for thinking about the data, thus facilitating idea exploration and theory validation (QSR, 1996). The subsequent coding structure is expressed in the form of a hierarchal tree diagram. Each node on the tree represents a thought, idea, or concept. All nodes are linked to data (contained in the document database) which are instances of the concept represented by that node.

### **Summary**

In keeping with the qualitative and quantitative nature of the research design, data analysis procedures used a range of qualitative and quantitative techniques also. These were selected on the basis of relevance and functionality for extrapolating findings pertinent to the proposed research questions and hypotheses. In particular, Rasch analysis, an innovative technique not previously applied to research of this nature, was found to be the most appropriate technique for analysing ordinal data. Its potential for separating items and persons on an interval continuum enables substantive conclusions to be formed, and facilitates additional statistical techniques. Results of Rasch analysis and other statistical procedures are presented in the following chapters.

## CONCLUSION

This investigation is set within the context of the Ericsson Cup, the premier domestic soccer competition in Australia. Specifically, the study set out to identify essential performance criteria for elite refereeing. Additionally, it attempted to assess the importance of these criteria, coupled with assessing the preparedness of officials in performing each criteria.

In conducting this research, a number of considerations needed to be addressed, particularly issues relating to data collection, i.e., the diversity of subjects (referees, referee assistants, referee inspectors, players and coaches), and the geographical locations in which these subjects resided (all states and territories except Tasmania and the Northern Territory). Accordingly, the research plan, design, and procedures, heeded such considerations. In doing so, a combination of qualitative and quantitative methods were employed to collect and analyse the data. Specially, the identification of refereeing performance criteria required the qualitative hybrid BARS technique to be employed. These criteria were used subsequently to assess referee performance quantitatively and qualitatively.

The distinct nature of the research design, plus its variant procedures, necessitated the integrity of the design to be assessed against quantitative and qualitative criteria. While not all these criteria were relevant to this investigation, the research design employed was shown to be appropriate for the nature of the investigation and the type of analyses undertaken in this study.

The diversity of research paradigms produced a variety of data, thus necessitating a range of data analysis techniques. These included the usual form of analysis applied to hybrid BARS data, i.e., cumulative percentage frequency analysis, plus additional parametric procedures such t-tests, analysis of variance (ANOVA and MANOVA) and regression analysis (Cook's Distance). Normally, these latter techniques are inappropriate for the ordinal data generated from BARS instruments. However, Rasch analysis is able to convert scores into interval measures, thus facilitating more detailed analysis. Lastly, unstructured qualitative responses were examined for common themes and issues using software (NUD•IST) specifically designed for exploring and indexing qualitative data.

The following chapters present the results of the data analysis. Specifically, Chapter 5 addresses issues concerning the suitability of BARS to identify and classify occupational competencies, including the suitability of this method with respect to other validity and reliability criteria not addressed within this chapter. Subsequent to this discussion, Chapters 6 and 7 examine results specifically relating to the importance of each criteria, plus the proficiency with which each criteria was performed.