

INTRODUCTION

This study evolved as a result of observing widespread concern among players, coaches, spectators and media about officiating (i.e., refereeing and umpiring) standards. Invariably, the concern pointed to officiating performances that were below expectations. However, the related discussion tended to be undertaken without a full understanding, and appreciation, of the officials' role.

This general lack of understanding occurs for two reasons. Firstly, the complexity of officiating makes the role difficult to describe in simple terms (Clegg & Thompson, 1993). Part of this problem stems from the diversity of sports that utilise officials, and the subsequent differentiation in officiating requirements for each sport. For example, the role of the official in contact sports has different expectations compared to the official in non-contact sports. However, organisations such as the National Officiating Program (an arm of the Australian Coaching Council) have made attempts to amalgamate the myriad of officiating competencies into general 'officiating principles' (National Officiating Program, 1997). Although the degree to which each principle applies to individual officials varies, these principles are designed to apply to all officials, irrespective of the sport with which they are involved.

The second reason relates to the lack of empirical research concerning refereeing performance. This point is reinforced by the small amount of specifically targeted research literature, and is somewhat surprising given the stated high importance of officials to sport (Evans, 1994; Spice & Thompson, 1994). Only the work of Anshel and Webb (1991) and Anshel (1995) have made attempts to clarify officiating requirements through recognised research techniques.

Given these issues, five research themes were identified to guide the present study investigating the performance requirements of elite soccer referees. These themes served to answer a variety of research questions and hypotheses. The first research theme was to identify performance criteria that are perceived by different stakeholder groups to be essential to elite soccer referees. The origins of this theme are grounded in Chapter 1. Here, evidence emerged that the role of a sports official is complex. In this chapter, the numerous roles and responsibilities officials assume are discussed. Further, peripheral issues that impact on officiating performance are explored. However, the literature indicated that specific requirements for officials within each sport are lacking. In combination, these issues provide a substantive basis for examining the nature of officiating work.

A technique for identifying occupational competencies, known as the Behaviourally Anchored Rating Scale (BARS), is examined in Chapter 2. The chapter reviews the

historical context of BARS, and examines the credentials of the technique to produce valid and reliable occupational criteria. Used initially to generate performance criteria for nurses (Smith & Kendall, 1963), the technique has undergone several modifications, and has been used subsequently to identify performance criteria across numerous occupations including basketball officiating, policing, university field instructors, and assistant nursing (Anshel, 1995; Bradley & Pursley, 1987; Conway, 1995; Stoskopf, Glik, Baker, Ciesla, & Cover, 1992). Inherently, the BARS process promotes behavioural descriptions of occupational requirements. However, as argued by contemporary sources (Hager & Beckett, 1995; Preston & Walker, 1993), this can lead to a narrow view of occupational performance. As a consequence, this study adopted a broader approach to competencies by including occupational knowledge, attitudes and abilities.

It is within this context that the performance criteria of elite soccer referees was addressed. In doing so, the present study was divided into two investigations, namely, the 'preliminary' and 'principal' studies. The preliminary study was designed primarily to identify the competencies required to referee the highest competitive level of soccer in Australia (Research Theme 1). Chapter 3 describes the design and procedures that were undertaken for this aspect of the study, and compares the results with other officiating research. Importantly, results from the preliminary study provide baseline data for the principal study, and allow the remaining four research themes to be addressed.

In Chapter 4, the research design and methodology is described and scrutinised (although the chapter focuses primarily on the principal study, some design aspects of the preliminary study are addressed also due to the close relationship between the two studies). As the principal study employed two research paradigms, an evaluation of the design (in quantitative and qualitative terms) is central to this chapter. Similarly, due to the unique focus of the study, and the development of a specialised research instrument predicated on findings from the preliminary study, the integrity of the research instrument is assessed in detail.

The next three chapters present results of the principal study. The first of these chapters, Chapter 5, assesses the suitability of hybrid BARS to identify and classify performance criteria (Research Theme 2). Given that the BARS variant used for this investigation was relatively new, fundamental assumptions which underpin the process required examination. In particular, analysis of the categorisation of competencies, into performance dimensions, forms a substantive part of the chapter. Additionally, general acceptance of the importance of competencies, from a wide cross section of soccer-stakeholders, is discussed.

The second results chapter, Chapter 6, presents finding from Research Themes 3 and 4. Theme 3 investigates the relative levels of *importance* and *preparedness* for each competency. From within these two perspectives, Rasch scaling enabled an unambiguous hierarchy of *importance* and *preparedness* to be determined for the first time. Subsequent findings and

discussion focus on significant differences found between the *importance* and *preparedness* of competencies, and the different perceptions among soccer stakeholders within the two perspectives (Research Theme 4).

As a means of verifying and expanding upon the findings presented in Chapter 6, results from the qualitative written comments are presented in Chapter 7. As these data were collected using an unstructured format, respondents were free to comment of any aspect of elite soccer refereeing (Research Theme 5). Not only did the written comments provide a mechanism to help elaborate on the findings identified in the quantitative data analysed in Chapter 6, they provided insight and direction into possible strategies that may improve the overall standard of elite soccer refereeing.

In the final chapter, limitations of the study are identified along with a summary of results as they relate to the five research themes that guided the investigation. Emerging from this overview are implications in relation to BARS methodology, referee assessment, and the training and development of referees. In conclusion, a number of future research initiatives generated by this investigation are presented.

CHAPTER ONE

REVIEW OF LITERATURE

Introduction

The continual growth of participation in sport has created an enormous demand for sports officials. This has resulted in players, coaches, and administrators, at all levels of sport, constantly seeking people who are competent at officiating (Clegg & Thompson, 1993). However, to be a competent official entails more than merely demonstrating an understanding of game rules (Brown, 1993).

From this context, the chapter provides an overview of the nature of the officiating role. In doing so, it examines three associated points, namely: major issues and concerns currently confronting the officiating profession; the performance criteria¹ essential for effective performance; and, the inter-relationship between these points. Although literature pertaining to officiating is not extensive, it does provide information which allows the diverse range of opinion and fact to be grouped into important themes, and the previous points to be explored. However, two important issues impact on this. Firstly, most literature has emanated from Britain and the United States. Secondly, the literature is predominantly non-specific to any one sport. As a consequence, issues and discussions raised in this chapter revolve essentially around this information (although soccer examples are provided where available and appropriate).

The chapter is divided into three sections. The first section defines officiating, and examines the plethora of roles and responsibilities assumed by officials. The second section analyses issues currently confronting the officiating profession, particularly those which impinge on the status of officials, i.e., bias, stress, and assessment. The final section provides an overview of relevant performance criteria associated with officiating, and, especially, how these factors relate to issues discussed in section two.

¹ Throughout this chapter and thesis, the term 'performance criteria' will be used on occasions as a collective noun for 'skills,' 'competencies' and 'attributes,' as these terms are used interchangeably in the literature.

DEFINING OFFICIALS: THEIR ROLES AND RESPONSIBILITIES

This section examines how officiating is currently defined, both in general and sport-specific terms. Fundamental to this discussion is an overview of the broad range of roles and responsibilities officials assume in the execution of their duties.

Defining Officiating

In its broadest sense, the term 'sports official' is an all inclusive designation applied across all sports. However, different sports tend to use specific titles for those people who perform officiating duties. For example, soccer, Touch (football) and basketball use the term *referee* to describe their officials (Anshel & Webb, 1991; Anshel & Weinberg, 1995; Ersing, 1992; Evans, 1994; Ferstle, 1978; Johnston & McNaughton, 1994; Kaissidis & Anshel, 1993; King, 1983; Peleg, 1991), while sports such as netball, Australian Football, baseball, softball and cricket call their officials *umpires* (Brown, 1993; Bruncker, Miran-Khan, & Carlisle, 1991; Koskelainen, 1997; Otago, Riley, & Forrest, 1994; Rainey, 1995; Sumner & Mobley, 1981). Interestingly, the Australian Coaching Council, which incorporates the National Officiating Program, also included *judges* – used in sports such as gymnastics and diving – under the officiating banner (National Officiating Program, 1997). However, the specific role of this latter group, i.e., assessing athlete performance, is fundamentally different from the arbitration role of officials from other sports. Thus, 'judges' are not considered within the context of this discussion.

A concise definition or description of the official's role is difficult to delineate. Numerous definitions of officiating, and the roles undertaken by officials, are presented in the literature. Some typical examples are provided below, starting with a rather sweeping and vague definition provided by Clegg and Thompson (1993). They contended that the fundamental aim of the official is "to let the game run smoothly within the rules" (p. 3). Although this description does give some indication of the official's role, it offers too much flexibility in interpretation. Other sources are similarly broad, although the perspectives they offer are varied. Grunski (1995) and Smith (1982) viewed officiating from an assertive perspective, describing the role as an "exercise in crisis containment" (Grunski, 1995, p. 39), or, to maintain "social order in tenuous social situations" (Smith, 1982, p. 33). However, not all descriptions have such dramatic overtones. Burke (1991) saw officials as arbiters, who provide judgements that are "within the spirit and intent of the rules" (p. 1). A final example sees officiating conceptualised as one of facilitation, with good officiating ensuring that the outcome of a game is dependent on the abilities of the players (Weinberg & Richardson, 1990).

Perhaps the most succinct definition is offered by Woodman (1995). He defined an official as “... any person who controls the actual play of a competition by the application of the rules and laws of the sport to make judgements on rule infringements, performance, time and score” (p. iii). Using this definition as a foundation, more specific and focused roles and responsibilities can be explored. For example, Weinberg and Richardson (1990, p. 4) provided explicit officiating responsibilities, namely to:

1. See that the event proceeds within the context of the rules.
2. Infer as little as possible.
3. Set and maintain an atmosphere for the enjoyment of the contest.
4. Show concern for the athletes.

However, underpinning all the descriptions of roles and responsibilities mentioned previously, the most recognisable role that the official assumes is to uphold the laws of the game (Clegg & Thompson, 1993; Ferstle, 1978; Johnston & McNaughton, 1994; Millane, 1988; Rains, 1984; Reilly, 1996; Wilson, 1986). All play in sports is governed by rules, with rules administered predominantly by international sporting federations. More often than not, the rules are explicit, with minimal latitude for interpretation. Nevertheless, in upholding the laws, the strict application of rules by officials may not necessarily be seen as ‘good’ officiating. It is maintained that officials should exercise some degree of discretion in their rulings (Reilly, 1996). This can be demonstrated through “minimal necessary intervention” (Rains, 1984, p. 152) by the official, rather than the enforcement of the rules through the active pursuit of rule breaches.

The diversity and complexity of the official’s role is compounded further by the requirement to control sporting contests which are often contextually based, yet framed by specific and objective rules which need to be applied without ‘fear or favour.’ This often presents contradictory responsibilities for the official, creating a variety of expectations.

Although the various descriptions, roles, and responsibilities of the official lack consistency, they nonetheless reinforce the importance of the officiating role. Officials have been described as the “essential third dimension” (Clegg & Thompson, 1993, p. 4) of sport – along with the players and coaches – and their contribution to the quality of the sporting contest is acknowledged in the sporting literature (Burke, 1991; Clegg & Thompson, 1993). This unique importance is magnified in comments by Evans (1994) and Spice and Thompson (1994). Evans contended that officials have been, and always will be, influential to sporting outcomes, while Spice and Thompson upheld the view that officials, through the singular position they hold within a game, can influence the quality of the sporting contest. While the extent of such influence is open to debate, it is interesting to note the existence of such an influence, particularly given that officials take no part in the playing aspects of sport.

Summary

In conclusion, the role of the official is one that cannot be ignored. Their responsibility as an arbiter, charged with facilitating successful play, is a concept that is broadly accepted in sport world-wide. The influence officials hold may vary from sport-to-sport (for example, referee input in golf is minimal as players tend to self-administer the rules), but very few – if any – sports are played without an official being present.

Most significantly, the role of the official is not static. As society changes, the role of the official is evolving also. With the advent of new technologies, and more knowledgeable players and spectators, there are a number of cognisant issues that have affected this evolution. These issues present further challenges to the officiating profession, and can have a marked impact on referees and refereeing standards. The following section addresses these issues, and provides a context for examining officiating competencies analysed in the final section of this chapter.

OFFICIATING ISSUES AND CHALLENGES

This section reviews seven issues which currently impact on the officiating profession, namely: the legal considerations of officiating; officiating training and education; the official's relationship with players and coaches; officiating bias; the assessment and evaluation of officials; officiating stress; and, the recruitment of officials. While these issues do not impact on the mechanical processes of officiating in an overt way, they are areas of concern that require deliberation by officials and respective sporting organisations. The first of these challenges, the ever increasing threat of litigation, is a relatively modern phenomena, but is one that is now linked inextricably to modern officiating.

Officials and the Law

As sport has become a more integral part of society, and a more professional approach to administration and playing has evolved, litigation is becoming more prominent. According to Patterson (1995, p. 1):

The relationship between sport and the law is now more evident than ever before, and all involved in the sporting industry, whether they be officials, coaches, managers, players, spectators, directors or administrators, can no longer ignore the principles of our legal system.

Civil actions in respect to sport-related injuries are the most common type of cases in sports law (Patterson, 1995). As such, it is not inconceivable that in fulfilling their responsibilities, individual officials may expose themselves to legal action. Specifically, sporting officials are legally, morally and ethically responsible for “exercising reasonable care to eliminate

foreseeable risks in the conduct of a game and to ensure that the game is conducted in a fair environment” (Fulluger, 1994, p. 13).

Numerous authors cited environmental conditions (in which the game is contested) as one area in which foreseeable risks may occur (Baer, 1992; Clegg & Thompson, 1993; Fulluger, 1994; Harvey, 1997; Millane, 1988; Murphy, 1995; Weinberg & Richardson, 1990). Examples of unsafe playing surfaces have been identified in courts of law where successful action by plaintiffs have been brought against officials for leaving broken glass on playing surfaces (Flygare, 1985), allowing athletes to use unsafe take-off boards (Weinberg & Richardson, 1990), and playing basketball on a wet floors (Murphy, 1995). Potholes, exposed wire, metal manhole covers, aluminium cans, wheel ruts and protruding pipes have also been discovered on playing surfaces. Given such circumstances, it is not improbable that serious player injury could result (Baer, 1992). As a consequence, if officials accept unsafe playing surfaces, and the playing surface leads to player injury, the official could be held accountable in a court of law, with accompanying financial and emotional consequences (Baer, 1992).

The provision of safe playing conditions may be imposed on the official by the rules of the sport (Millane, 1988). As an example, such requirements can be found within the laws of soccer. The referee has discretionary power to suspend or terminate a game whenever deemed necessary under the internationally recognised Laws of Soccer (Soccer Australia, 1998). The jurisdiction also allows the referee similar powers to deal with additional safety issues such as spectator interference, the condition of fixtures and equipment, and player injury (Soccer Australia, 1998). Kuhn (1997) concluded that where a rule is in some way associated with the safe playing of a sport, breach of this rule is more likely to, and should, attract liability for those responsible for the rules’ implementation.

The types of additional safety responsibilities imposed on soccer referees was given currency in the professional literature (Baer, 1992; Karns, 1986; Kuhn, 1997; Murphy, 1995; Weinberg & Richardson, 1990). These writers supported the need for officials to take far-reaching responsibilities for safety issues. This being the case, officials can be liable for failure to: remove a participant who persistently breaches the rules or uses excessive violence; enforce strictly the safety rules designed to prevent injuries; stop an event when the safety of participants is threatened by spectator violence; and, enforce safety guidelines, including policies involving blood and infectious diseases (Murphy, 1995). Although many of these situations may be deemed minor, irregular, or not under the jurisdiction of officials, officials should be aware of potential litigious actions, and be prudent in the administration of their responsibilities. Only recently, an English court found a rugby referee liable for injuries sustained to a player when a scrum collapsed (Stoner, 1998). The judge found there were approximately 25 scrum collapses during the match, a figure the court accepted as

excessively high (evidence suggested six collapses per game more usual). As a consequence, the referee was deemed to have breached his duty of care to players in failing to maintain acceptable standards of scrummaging.

The litigious circumstances discussed thus far emanated essentially from the *inactivity* of officials. However, two cases have demonstrated that the *actions* of officials can lead also to legal action. The first case involved two officials who attempted to administer first-aid to a football player who had injured his neck and back (Weinberg & Richardson, 1990). A lawsuit was brought against the two officials for assisting in administering first-aid when they were not qualified to do so, and thus aggravating the injuries. Although officials are responsible to injured athletes (Karns, 1986), Weinberg and Richardson (1990) maintained that an official's response to player injury should be to bring the injury to the attention of others (team trainers and the like), and not the provision of first-aid treatment.

In the second case (cited by Flygare, 1985), a basketball official called a foul against a player in a team from the University of Iowa. The resultant free throws to Purdue University led to Purdue University winning the game. A retail outlet in Iowa City – Hawkeye John's Trading Post – filed a claim against the official for \$175 000, plus exemplary damages. It was claimed that the official's decision constituted professional malpractice, which led to the loss of potential market for the products in the store that proclaimed the University of Iowa the champions of the Big Ten conference. The owners of the store also asserted that they had suffered emotional stress, anxiety, and loss of business goodwill because the official's decision was "baneful, outrageous, and done with heedless disregard for their rights" (Flygare, 1985, p. 581). While the court dismissed the claim, this situation is symptomatic of the scrutiny officiating decisions can attract in a litigious society.

Notwithstanding the problematic nature and potential legal pitfalls of officiating liability, Patterson (1995) maintained that the prospect of liability should not be seen as an unwarranted intrusion into the world of sport. He concluded that the prospect of liability encourages a safer sporting environment, and that "liability may be seen as a legal device promoting the safe administration and practice of all sports" (Patterson, 1995, p. 1). Given these comments by Patterson, officials need to see merit in what he has to say, and undertake a proactive role in the provision of safe sport. To aid in this provision, legal implications pertaining to officiating should be incorporated as a major component of officiating training programs (Davis, 1990; Murphy, 1995; National Officiating Program, 1997). Education in this area of responsibility, in conjunction with the many other aspects of refereeing performance, has the potential to minimise legal actions against officials.

The Training and Development of Officials

The training and development of sports officials has traditionally involved training in “knowledge of the rules” (White, 1991, p. 24). This is based on the premise that a good official needs to have a thorough understanding of the laws which govern play. However, it is now broadly documented that the training and development of good officials is far more intricate than the acquisition of rule knowledge *per se* (Bies, 1990; Davis, 1990; Kaissidis & Anshel, 1993; O'Connor, 1994; Peleg, 1991; Taylor, Daniel, Leith, & Burke, 1990; White, 1991).

To illustrate this point, White (1991) developed a simple model to depict the qualities required to officiate netball (see Figure 1.1). Although White acknowledged the need for netball umpires to have a sound knowledge of netball rules, she emphasised the necessity for umpires to use other skills, such as communication, observation, and movement and positioning, in order to apply game rules. These qualities were seen as equally, if not more, essential than mere rule knowledge in developing the “complete umpire” (White, 1991, p. 24). Importantly, such a model would be similarly applicable to other sports.

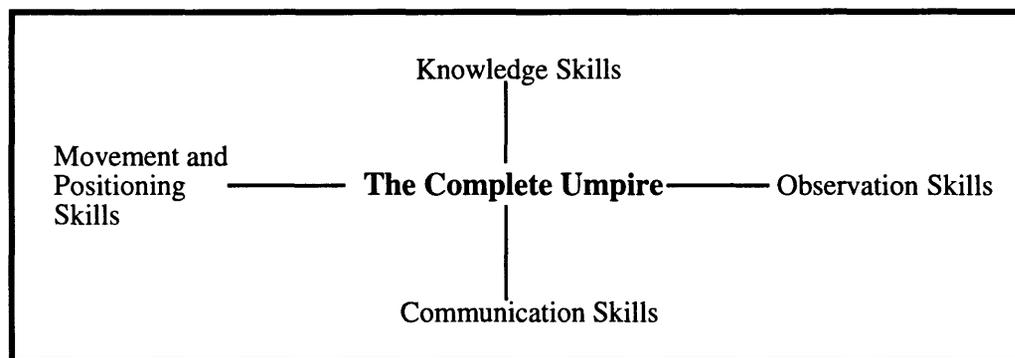


Figure 1.1 Skills required for netball officiating

(adapted from White, 1991, p. 24)

The officiating qualities that White (1991) alluded to in Figure 1.1 are supported by other sources (Anshel, 1989; Australian Coaching Council, 1996; Evans, 1994; Soccer Australia, 1996; Weinberg & Richardson, 1990). By implication, therefore, these qualities should become the foci of officiating development programs. However, in the development of such programs, all relevant and applicable skills need consideration for inclusion in training and development programs (see Table 1.1).

Although the list of qualities and attributes supplied in Table 1.1 is extensive, the comprehensive development of these competencies should not be seen as an immediate prerequisite for an officiating career. As Bies (1990) stated, an official’s education “is an

Table 1.1 Types of officiating skills by author

Officiating Skill	Author
stress management	Kaissidis and Anshel (1993), National Officiating Program (1997), Rainey (1995), Taylor et al. (1990)
assertiveness	Taylor et al. (1990)
confidence	Rainey, Larsen and Stephenson (1989a), Taylor et al. (1990)
physical preparation	National Officiating Program (1997), O'Connor (1994), Peleg (1991), Soccer Australia (1995)
nutrition	Peleg (1991)
psychological preparation	Davis (1990), Peleg (1991), Taylor et al. (1990)
concentration/attention skills	Murphy, Tanzor and Trotter (1995)
decision-making	Murphy et al. (1995), National Officiating Program (1997)
legal responsibilities	Davis (1990), Murphy et al. (1995), National Officiating Program (1997)
self reflection skills	National Officiating Program (1997)

ongoing process whereby one continues to gather more information and experiences through advanced and specialized training” (p. 64).

In an attempt to draw together the myriad of performance criteria and ideas that comprise the preparation of sports officials, the Australian Sports Commission has developed an ‘officiating general principles’ course (National Officiating Program, 1997). This course endeavours to provide sports officials, across all sports, with the knowledge and skills that underpin the officiating process. It included modules related to: legal responsibilities and risk management; self-reflection; communication skills; dealing with conflict; and, fitness and recovery techniques (National Officiating Program, 1997). This course now supplements sport-specific courses conducted by state and national sporting organisations, and can lead to formal recognition by the National Officiating Accreditation Scheme (National Officiating Program, 1997).

Although individual sporting organisations do, and will continue to, train and develop officials through formalised curriculum and instruction (for example, courses offered through the National Officiating Program), officials also learn about their profession through a ‘hidden curriculum.’ This concept is depicted in research conducted on baseball umpires (Rainey et al., 1989a). Informal comments from these umpires indicated that they learn how to project confidence in their decisions (even if they are unsure of their decision) in order to maintain control of the game. Obviously, such skills would be difficult to formalise in any educative setting. However, the ability to ‘sell’ decisions represents an important strategy in the official’s repertoire of skills (Kaissidis-Rodafinos, Anshel, & Sideridis, 1998).

On the surface, it appears to be difficult to reach a general conclusion on the best approach to train officials, or which competencies require the most emphasis. The diversity of sports, coupled with the varying standards and officiating roles within each sport, conspire to make definitive judgements difficult. This problem is further compounded by sentiments expressed by Rokosz (1989), who surmised, “you can’t do much with people who don’t have inherent talent for the job” (p. 2). In light of these concerns, perhaps officiating training and development needs to be contextual, and, from this platform, education programs can be developed which meet the needs of each sport and each official.

However, notwithstanding the sport involved, or the sport’s competitive level, officials operate within an interactive work environment. In addressing this point, officiating training programs address the symbiotic relationship within officiating teams, e.g., how a soccer referee interacts with his/her two assistant referees to facilitate the optimal administration of game rules. Nevertheless, the interactive environment of the officials work extends further, and encompasses interaction with players and coaches also. Fundamentally, players and coaches are ‘clients’ of officiating decisions. Yet the nature of the relationship between these groups, and how to promote positive interaction, is given scant attention in current officiating education programs. As can be seen from the following discussion, the need for officials to comprehend the nature of this relationship is central to their effectiveness, and to the smooth progress of a game.

The Relationship of Officials with Players and Coaches

An uneasy relationship tends to exist between game participants (players and coaches) and sports officials (Burke, 1991; Gerisch & Kosel, 1988; King, 1983; National Officiating Program, 1994; Phillips, 1985; Rainey, 1994; Rainey, Schweickert, Granito, & Pullella, 1990; Shults, 1991). Historically, game participants and officials have maintained an adversarial relationship (Martin, 1996), with manifest preconceived notions about each other’s role and relevance. Typically, game participants view officials as a necessary evil in sport, while many officials hold negative stereotypes of coaches, and can be cynical of players’ actions. This view, according to Burke (1991), is maintained because most interaction is of an aversive or unpleasant nature.

The attitudes held by officials to game participants could be explained by the type of treatment officials receive from time-to-time. In a study of collegiate basketball officials in America, Burke (1991) cited a variety of concerns about the officials’ relationship with coaches. These concerns were: the intentional baiting by coaches; officials having to deal with coaches who do not understand the rules; coaches influencing the selection and retention of officials; the performance ratings of officials by coaches; and, coaches’ criticism of officials in the media. According to Burke (1991), this situation is exacerbated by officials not being able to respond in kind.

Most alarmingly, three studies have shown that coaches and players have assaulted officials in a number of sports and across a variety of cultures (Rainey, 1994; Rainey & Duggan, 1998; Rainey & Hardy, 1999). The sports investigated included softball and baseball in the USA (Rainey, 1994), basketball in the USA, (Rainey & Duggan, 1998) and rugby in Great Britain (Rainey & Hardy, 1999). Moreover, results from these studies indicated the types of assaults were diverse. Although many assaults were of a minor nature, e.g., pushing and shoving, a marked proportion were of a more serious nature, e.g. punching, kicking, and choking. One consequence of such incidents is the increased use of the legal system, by officials, seeking recourse against aggressive actions (for examples, see Cross, 1998).

Exacerbating this animosity is an *expectation* that players and coaches be hostile towards sports officials. In a survey of Major League Baseball supporters, Rainey et al. (1990) indicated 62% of male respondents thought that it was acceptable for players and coaches to “argue heatedly” (p. 125) with umpires, while 18% thought it was acceptable to “yell and swear” (p. 125) at the umpire. Rainey et al. concluded that, in large crowds, thousands of supporters accept players and coaches who are verbally hostile to officials. The type of ambivalence towards antisocial behaviour, reported by Rainey et al., was referred to as “sanctioned violence” by Phillips (1985, p. 6). It now appears to be more accepted and common-place in modern sport, and is reflected in comments by Smith (1982, p. 36), who stated:

In most everyday life situations, no matter how angry or upset you are with someone, the mores of social responsibility require that you show some sort of consideration for the feelings of the person with whom you are dealing and somehow attempt to negotiate your differences. No such rules seemed to apply to relationships with a referee.

Central to the relationship which exists between game participants and officials is the dichotomous roles these subgroups assume within the sporting context (Gerisch & Kosel, 1988; Phillips, 1985). Players and coaches are most concerned with the outcome of the game (Alker, Straub, & Leary, 1973), and, consequently, take actions that are likely to meet this outcome (Phillips, 1985). Continuing this theme, it has been maintained that the role of the official is perceived by some participants as “contrary to their [the players and coaches] best interests” (Anshel, 1989, p. 32). This notion is demonstrated clearly in soccer, where, in order to gain an advantage over their opponents, players are now taught how to break the rules without getting caught (Shults, 1991). The recent 1998 Soccer World Cup illustrated this point. Here, many instances occurred where players deliberately tried to deceive the referee in order to obtain an advantage over their opponents (Elleray, 1998). In contrast to such motivations, officials are responsible for the fair and proper conduct of the game. Additionally, they are presumed to be indifferent to the two competing teams (Cross, 1998), and, as such, should have little or no concern for the eventual outcome of the game.

Such variance in attitudes was initially brought into sharp focus by Gerisch and Kosel (1988). Their research, conducted within the context of professional German soccer, reported that referees' opinions pertaining to aggressive play differed considerably from those of players and coaches. The players and coaches did not view provocation and threats of foul play as aggressive behaviour. Furthermore, players were more likely to see fouling as an acceptable form of 'discouraging' the opposition, while coaches were more likely [than referees] to condone aggressive behaviour, particularly if the behaviour aided in gaining an advantage over the opposition.

Supporting the findings of Gerisch and Kosel, Shults (1991) claimed that soccer players now possess an attitude where it is the responsibility of the referee to catch players committing fouls. In doing this, Shults (1991) maintained that players were absolving themselves from ethical conduct. More pointedly, player attitudes gave rise to the intentional foul being viewed as a legitimate strategy (Shults, 1991), and, accordingly, the professional foul and subsequent faking of injury has become common place in soccer (King, 1983; Shults, 1991). These attitudes and strategies have been interpreted as playing against the referee, rather than playing against the opposition (King, 1983; Knox, 1998). Ersing (1992, p. 27) concluded that:

... the attitude of the player toward self discipline and the attitude of the coach towards rules of decency and fair play cannot be ignored. They are critical variables that will determine the level of fair play and, in the final analysis, the quality of soccer in a match.

Aside from the general antagonism and distrust that exist between game participants and officials, these groups have many similarities and should have a close working relationship. These similarities (adapted from Shults, 1991, pp. 17-20) can be viewed as common threads which provide strong links between the respective roles of officials and game participants. The only significant difference between these groups is that each sub-group is responsible for different aspects of the game. A summary of the similarities between officials, players, and coaches is provided below:

1. **Respect for the game:** players, coaches and officials are involved because they love the game (very few players, coaches and officials are full-time and, as a consequence, need to hold down other jobs). Players, coaches and officials can take equal pride in the success of a match.
2. **Know when to quit:** when players, coaches and officials no longer look forward to a game they should quit. Players and coaches seem to lose the thrill of winning, and begin to fear losing. Moreover, players tend to lose their enthusiasm for training and playing becomes a chore. Officials strain to maintain fitness and begin to dislike

players. When coaches and officials feel themselves constantly angry and continually in need of defending themselves, it is time to quit.

3. Feel for the game: it is the role of players, coaches and officials to orchestrate a fair game. Further to this, they all have the responsibility to keep the game under control.

Additional similarities, specifically between coaches and officials, have also been identified (Shults, 1991, pp. 17-20). These were:

1. Being most effective when hardly noticed: the game is most successful when the game centres around the players. Coaches need to avoid over-coaching, while officials need to avoid over-officiating.
2. Assuming positions of authority: coaches and officials represent authority to players as both have some degree of control over the players and their actions.
3. Assuming teaching roles: coaches teach strategy and skills, while officials teach the rules of the game and fair play.
4. Taking responsibility for player health and safety: coaches and officials are concerned for the health and safety of players. It is asserted that coaches must not teach or sanction unfair and 'dirty' tactics, and, as noted previously in this section, the official has responsibility for a range of issues which impact on player safety.

Despite these similarities noted by Shults (1991), it is probable that disagreement between officials and coaches/players, particularly in combative games which require subjective rulings, will remain. Limitations exist in the conduct of games which would seem to limit the possibility of resolution. Examples include the low scoring nature of some games (such as soccer, where relatively minor incidents and decisions can take on great importance, see Auer, 1985), and the impossible expectation that officials will be 100% correct in their decisions (even with the advent of video-replays at the elite level of particular sports, some officiating decisions still result in controversy). While such limitations exist, officiating decisions will continue to cause debate, and division, between officials and game participants, even leading to allegations of officiating bias and favouritism. Such allegations, while difficult to sustain, are very real and can threaten the integrity of the sport, and the status of the official.

Officiating Bias

The decisions that sport officials make are supposed to be governed by a formal set of rules. Decisions should be based on objective evidence and free of any bias or emotion which can pervade a game (Burke, 1991). An official is, in the simplest sense, required to make

judgements and decisions that are based on what they see in the sporting contest. However, despite the official being appointed to be an impartial arbiter, a number of authors contended that many officiating decisions are made due to the social context of the game, or because of psychological and political considerations. As such, decisions are not based solely on the actions of players (Ansorge, Scheer, Laub, & Howard, 1978; Askins, Carter, & Wood, 1981; Lehman & Reifman, 1988; Rainey et al., 1989a; Wanderer, 1987), but may be attributed to officiating incompetence or officiating bias (Ansorge et al., 1978).

In an attempt to conceptualise the context of an official's decision-making, Lehman and Reifman (1988) explained that officials do not make decisions in a social vacuum. They sometimes work in a social environment that can often contain large and vocal crowds, which may have a direct impact on officials (Lehman & Reifman, 1988; Mitchell, Leonard, & Schmitt, 1982). The public nature of their work, coupled with the ambiguity of events that officials need to adjudicate, prevent officials from being oblivious to crowd reaction (Askins et al., 1981; Phillips, 1985).

This context is supported by research which examined professional soccer in England and Scotland. It showed that the number of 'player expulsions' and 'penalties scored from' increased in a linear trend based on mean crowd size (Nevill, Newell, & Gale, 1996). The effect was further illustrated by Pollard (1986), who showed that the 'home advantage' in English soccer disappears when crowds do not favour any one team (for example, in 'local derbies,' i.e., games played between teams in the same city and access to grounds is equal for both sets of supporters).

The concept of 'home advantage,' in relation to the impact of the crowd on officiating decisions, is worthy of examination. Courneya and Carron (1992) defined the home advantage as "the term used to describe the consistent finding that home teams in sport competition win over 50% of the games played under a balanced home and away schedule" (p. 13). Home teams have won significantly more games in soccer (Glamser, 1990; Nevill et al., 1996), cricket (Sumner & Mobley, 1981) and basketball (Varca, 1980). As a result, the home advantage is seen as a good predictor of game outcome in professional sport (Courneya & Carron, 1992; Dockery, 1997; Snyder & Purdy, 1985), and is ranked alongside 'team quality' in importance for the pursuit of team outcomes, i.e., winning.

When the home advantage is examined in the context of sports officials, Glamser's (1990) research found that referees in English professional soccer rarely have an affiliation with the home team. However, decisions which have a negative impact on the home team are negatively reinforced by the crowd. Thus, honest mistakes are noticed, and the official will be tempted to 'even things out.' Similarly, in general terms, it has been hypothesised that officials make more subjective judgements in favour of the home team (Nevill et al., 1996).

Not only has 'playing at home' indicated officiating bias in terms of winning and losing, but also in the allocation of fouls and 'major decisions.' It is reported that players are twice as likely to receive an official caution in soccer when playing away from home (Glamser, 1990). In basketball, referees may be intimidated by home audience harassment, and therefore, call less fouls against home teams (Varca, 1980). Similarly, Jurkovic (cited in Nevill et al., 1996, p. 185) found that players "tried to get away with more" in home games because they perceived officials were more intimidated by crowds to award more calls to the home team. Observations akin to those by Glasmer and Jurovac have also been reported in Test-match cricket, where visiting batsmen are more likely to be given out LBW (leg before wicket) in India than in any other Test playing country (Sumner & Mobley, 1981). It would appear that players are able to sense an 'unnatural' benefit when playing home games, and, consequently, take advantage of this benefit to meet their own ends.

Despite this evidence, the home advantage, when measured against win-loss ratios and foul frequencies, does not illustrate the full picture of officiating bias. Additional evidence to support officiating bias has focused on the treatment of 'star' and 'stand out' players. It was hypothesised by Lehman and Reifman (1988) that evidence based on home-and-away fouls may not necessarily indicate referee bias. They maintained it may merely indicate that home teams commit less fouls, or the home crowd may simply be able to heighten the visiting players' arousal and anxiety levels, thus resulting in more reckless and dysfunctional behaviour (Nevill et al., 1996; Varca, 1980). To control for these factors, Lehman and Reifman (1988) examined American professional basketball for any interaction between the site-of-game and player status. They deduced that 'star players' (those players who had participated in NBA All-Star games) were called for significantly fewer fouls ($p < .05$) at home games, while non-star players were *not* called for significantly fewer fouls when playing at home. This research indicated that officials may not apply the same standards to all players, irrespective of where the game is played. Comparable officiating bias towards 'star players,' both for and against, has also been reported in baseball (Rainey, Larsen, & Stephenson, 1989b) and gymnastics (Ansorge et al., 1978; Scheer & Ansorge, 1975; Scheer, Ansorge, & Howard, 1983).

Similarly, players who 'stand out' from other players, e.g., players with red hair, a beard, differing skin colour, or exceptionally long hair, may be subjected to less favourable forms of bias. In supporting research, Glamser (1990) investigated the effect of game location on player misconduct for a professional London soccer team by reviewing the location and frequency of official cautions issued over a seven-month period. It was concluded that black players were cautioned more often than white players, particularly in out-of-London away games. Glamser (p. 47) stated:

No doubt that blacks were playing more aggressively in response to a hostile crowd, but the referees had to be watching them closely as well, creating a self-fulfilling prophecy situation.

Glamser did not directly accuse the referees of racism, but rather inferred that referees often connected various fouls with players who possess distinguishing physical characteristics. Officials are likely to connect subconsciously such players more readily with fouls that they commit, and therefore the second or third foul is perceived as such. For more non-descript players, the second or third foul may be perceived as their first foul only.

Although the evidence presented thus far appears to support the existence of officiating bias, obtaining empirical evidence on officiating bias *per se* is difficult and almost impossible to substantiate. This is because numerous other reasons can be offered for discrepancies in match results and player behaviour. These include: familiarity with the playing surface and arena; fatigue of the visiting team; social support for the visiting team; displaced aggression of the visiting team; and, intimidation of players from the visiting team (Glamser, 1990). Moreover, it is maintained that losing can bring into question team objectives, task performance, and perceived superiority of individual and team skills (Ross, 1997). Consequently, when such questions arise, team cohesion is threatened. Ross (1997) continued to assert that it is easier for players and coaches to externalise problems, i.e., blame the referee, than to look within oneself to the true causes of defeat. Obviously, when the full impact of other game-related factors are assessed, the ramifications of officiating bias, if it does exist, are difficult to isolate from several other manifest factors within a sporting contest. More research may help to explain and clarify the phenomena of officiating bias (Rainey et al., 1989b).

Nevertheless, officials need to guard their reputation from bias at all times. The public nature of officiating places officials under intense scrutiny. Media, players, coaches, and spectators all debate the performance of sports officials, in some way, from time-to-time. If, as the literature indicated, officiating bias is such a major issue, officials need to be aware of any perception of bias and have strategies which down-play or avoid the factors that contribute to such perception. Even when a mistake is made, it is important that it is seen as an honest mistake (Wilson, 1986).

Such impartiality is central to maintaining the integrity of the official's reputation and status. This is particularly important if the official is to be judged or assessed in any formal, or informal, manner. As indicated previously, the performances of officials are judged from many different perspectives, and it is important that officials minimise the perception that extraneous factors have impacted on their performance. If such perceptions do exist, fair and equitable judgements of officiating performance are difficult and can result in an official's performance being judged unfairly.

Assessment and Evaluation of Officials

As noted, officiating decisions are made in a variety of contexts, often under significant time-pressures in fast-moving, competitive, and physically vigorous games. At the same time, there are expectations from all observers, that decisions are objective and accurate. When viewed from these perspectives, the *work* of the official is somewhat problematic (Rains, 1984). Conversely, the *evaluation* of an official's performance is less problematic, with decisions, particularly in elite sport, validated or rebuked by numerous video replays.

Added to the complexity of this problem are two additional issues. Firstly, officials are not expected to call every rule infraction (Rains, 1984). If officials were to call every infringement, many games would continually stop-and-start, with players and spectators deriving less enjoyment than would normally be the case. Secondly, different officials have different methods of being effective (Australian Coaching Council, 1996). Officials, just like players, have different officiating 'styles,' often as a manifestation of differing personalities (Clegg & Thompson, 1993). The variation in the interpretation of officiating effectiveness, which these factors alone can cause, magnifies the assessment and evaluation conundrum. However, notwithstanding these problems, numerous sports are now attempting to evaluate formally the performances of their respective officials (Askins, 1987; Australian Coaching Council, 1996; National Officiating Program, 1994; Rains, 1984; Stocklin, 1988).

To aid in the evaluation process, Stocklin (1988) developed an evaluation framework (see Figure 1.2). Amongst a number of variables and inputs, this framework encompassed the following salient points: multiple methods of rating (including officials self rating); the final rating being determined by more than one person (the rating committee); and, a necessity for raters to be reviewed periodically.

The need for multiple forms of rating, as advocated by Stocklin (1988), can incorporate self evaluation, raters' assessment, and formalised test scores. This latter form, while valuable, does not necessarily supply information related to the quality of officiating performance. However, evaluations conducted by officials (i.e., self-assessment), independent 'third parties', and game participants, can provide a clearer (although still subjective) indication of the quality of an official's performance.

Conversely, the value of self-assessment is well recognised and encouraged (Askins, 1987; Australian Coaching Council, 1996). Particularly, it is recommended that self-assessment be made in the context of three key inter-related areas, namely: game control; decision-making; and, communication (Australian Coaching Council, 1996). The process is seen as an important aid in the improvement of officiating effectiveness (Australian Coaching Council, 1996), as it allows the official to recognise past mistakes and the situations where mistakes occurred (Askins, 1987).

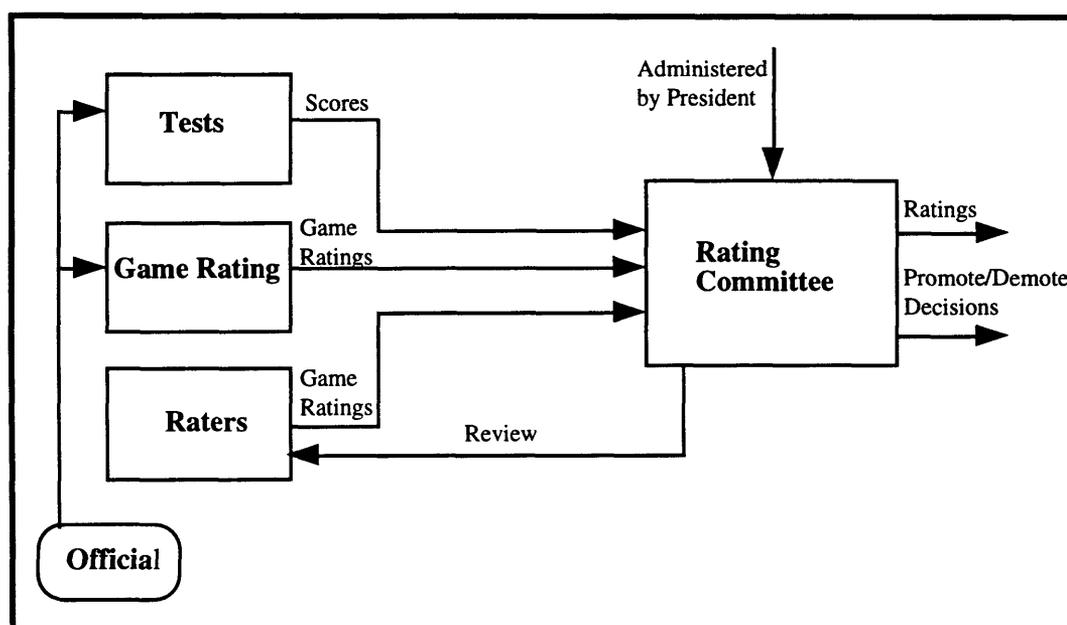


Figure 1.2 Rating system structure

(adapted from Stocklin, 1988, p. 32)

Despite the inherent benefits and advantages that can be derived from self-assessment, Baer (1991) claimed that this assessment method should occur in conjunction with third-party assessment. The use of officiating inspectors (also referred to as supervisors, reviewers, or assessors) is identified as the main form of third-party assessment. Importantly, it is maintained that most officials see this as the most effective means of targeting areas of strength, as well as those needing improvement, because inspectors ideally offer independent and unbiased views (Baer, 1991).

This view, while logical in theory, was challenged by Ersing (1992). He disclosed that inspectors did not always provide objective assessment. From the 1990 Soccer World Cup, Ersing obtained evaluation data about referee performance from referee assessors (inspectors). He found that the control and management of inappropriate player behaviour – through the use of cards (official cautions and send-offs) – was generally perceived by inspectors as a weakness in refereeing performance. Ersing inferred that such a judgement on the referee was unfair, as the attitudes of the players towards their own behaviour, and the attitudes of the coach towards the rules and fair play, were factors that should not have been ignored. Ersing concluded that “all too often, assessors place disproportionate responsibility on the referee to control individual team behaviour” (p. 27). The implication of these findings is that inspectors allow their individual biases and prejudices to dictate assessment ratings of referees. Logically, such circumstances are not objective, and call into question the value of third-party assessments.

However, the use of inspectors is only one form of third-party assessment. The utilisation of coaches and players, to conduct assessments, also constitutes a valid form of assessment. Since these people observe officiating performance from close proximity, they are in a unique position to assess performance. Nevertheless, their use as assessors is seen as problematic by a number of sources (Baer, 1991; Burke, 1991; Schwartz, 1977; Wilson, 1986). It is argued that coaches are not entirely objective in their assessment of an official's performance and, subsequently, not always qualified to rate officials (Schwartz, 1977). Additionally, the assessment provided by coaches may be inconsistent. For example, Soccer Australia have found that referee evaluations, submitted by Ericsson Cup clubs (but invariably completed by coaches), can be notably contradictory about the same refereeing performance (Power, G. 1997, pers. comm., 13 June). This reflected previous statements by Straub, who concluded that "... apparently, coaches and fellow officials evaluate officiating ability using different criteria" (cited Schwartz, 1977, p. 19).

Perhaps because of these reasons, one of the most contentious issues, for officials, is coaches continually judging and evaluating an official's performance. This is particularly true given that officials do not get the opportunity (either through formal evaluation processes or the media) to evaluate coaching performance (Burke, 1991). This situation is intensified by coaches who, when evaluating officiating performance, lack a clear understanding about the intricacies of officiating (Burke, 1991). Moreover, it is claimed coaches can be influenced by crucial decisions, rather than judging officiating performance from a broader perspective (Burke, 1991). The preoccupation by coaches on isolated incidents, and the impact of such incidents, is questioned by Auer (1985), who asked, "How many referees' errors wind up in the goal? Very, very few. Yet ... coaches keep screaming at the referee" (p. 30).

Despite the ambivalence expressed by these sources about the effectiveness of such evaluations, it is argued that constructive comments from coaches can provide "valuable gauges for improving performance" (National Officiating Program, 1994, p. 14). Similar sentiments have also been expressed towards player evaluations being used to assess officiating performance (Brown, 1993; Spice & Thompson, 1994). However, the use of players has also attracted a degree of scepticism. Wilson (1986) warned that many [soccer] players, including top-class players, only have a basic knowledge and understanding of the laws of soccer. Wilson argued that this can lead to unwarranted complaints by the players about a referee's decision. Similarly, player comments should not be seen as definitive, as players often complain simply because things are not going their way (Baer, 1991). The implications of the argument mounted by Wilson and Baer is that players are not the most appropriate source to evaluate objectively an official's performance.

There would seem to be no definitive method for evaluating officials which meets with 100% approval from all stakeholders. Perhaps the problem lies in the rater not being exactly certain of the criteria against which they are rating the official. As Clegg and Thompson (1993) noted, 100 officiating experts are likely to give you 100 different listings of the essential qualities of a master official. However, no matter which methods or criteria are used to conduct the evaluation, Stocklin (1988) emphasised the importance of the ability of those administering the evaluation, and the need for raters to apply assessment criteria in a consistent manner.

Officiating Stress

Although the assessment and evaluation of officiating performance is seen as desirable and important, it is also cited as a major cause of stress for officials (Taylor & Daniel, 1988; Taylor et al., 1990; Zoller, 1985). Stress, defined as the sudden and short-term exposure to demanding situations that exceeds the individuals resources (Anshel, 1990), is inherent in sporting competitions, and its effect on the official can be substantial (Goldsmith & Williams, 1992). It is reported that officials now suffer from many stress-related illnesses, including heart disease, peptic ulcers, high blood pressure, headaches, skin disorders, colitis, diarrhoea and cancer (Goldsmith & Williams, 1992; Weinberg & Richardson, 1990; Zoller, 1985). The following discussion examines the nature, causes and dimensions of officiating stress, with particular reference to the roles played by players and coaches in its occurrence.

With the previously noted physical problems becoming manifest, stress is now recognised as a major officiating issue (Anshel & Weinberg, 1995; Goldsmith & Williams, 1992; Rainey, 1995; Taylor & Daniel, 1988; Zoller, 1985). Consequently, officiating stress is attracting increased research across a variety of sports, including: basketball (Anshel & Weinberg, 1995; Kaissidis & Anshel, 1993; Kaissidis-Rodafinos et al., 1998; Purdy & Snyder, 1985; Rainey & Winterich, 1995); soccer (Taylor & Daniel, 1988; Taylor et al., 1990); American football (Goldsmith & Williams, 1992); volleyball (Goldsmith & Williams, 1992; Stewart & Ellery, 1998; Stewart & Ellery, 1996) baseball and softball (Rainey, 1995); rugby (Rainey & Hardy, 1997) and, wrestling (Smith, 1982). Moreover, the research has examined the effects of stress across a variety of officiating levels, ranging from 14 year old hockey officials (Spice & Thompson, 1994) to accredited international soccer referees (Taylor & Daniel, 1988).

Table 1.2 summarises the *general* causes of officiating stress. It should be noted from this table that not all causes of stress are related to on-field performance. A number of causes are specifically related to time pressures imposed on the official and the unpopular status of the officials' role.

Table 1.2 General sources of officiating stress

Source of Stress	Author
The number of games that an official is required to preside over	Zoller (1985)
The continual review process	Taylor and Daniel (1990), Taylor et al. (1988); Zoller (1985)
Travel commitments	Purdy and Snyder (1985), Zoller (1985)
The unpopular role of officiating	Purdy and Snyder (1985), Smith (1982)
Inadequacy of performance	Rainey (1995)
Pressure from the crowd	Zoller (1985)
The avocational nature of the officiating role	Goldsmith and Williams (1992), Purdy and Snyder (1985), Rainey (1995), Zoller (1985)

In addition to these general causes, a number of sport-specific studies have detailed stress-related situations within given sports. Research by Taylor and Daniel (1988) identified a number of situations and events that contribute to the psycho-social stress in soccer officiating. The four most stressful circumstances were identified as: having a bad game; working with poor officials; making a bad call; and, players and coaches who do not understand the laws (p. 540). Subsequent factor analysis of response items determined six over-riding factors that contribute to officiating stress. These factors were: interpersonal conflicts; fear of physical harm; time pressures; peer conflicts; role-culture conflict; and, fear of failure (Taylor & Daniel, 1988, p. 542).

Similar findings were also reported for American, Greek, and Australian basketball referees. They disclosed a number of stressful events, including: making a wrong call; verbal abuse by coaches; threats of physical abuse; being in the wrong location to make a call; and, experiencing an injury (Kaissidis & Anshel, 1993; Kaissidis-Rodafinos et al., 1998). Specific examples encompassed by these events were: use of profanity by the coach; threat to strike [hit]; and, not feeling well.

A consistent thread through all these studies is the threat of verbal and physical abuse. The previously mentioned study by Rainey (1994) cited a range of physical abuse inflicted on officials, including spitting, choking, shoving, and grabbing. It seems that such abuse is not in decline. Further instances of biting, punching, and choking have been recently reported (Cohen, 1997). Although these forms of abuse appear to be prevalent across numerous sports, Knox (1998) saw professional soccer as being worse than most sports. Particularly, Knox cited the continual crowding and pushing of the referee as becoming endemic in professional soccer. He surmised that “only disturbed masochists” (p. 50) would continue to volunteer for this type of ritual abuse.

The impact of these stressful situations will vary from official to official. Individual officials have their own mechanisms to cope with stress, some being more effective than others. Nonetheless, as a general guide, it is hypothesised that the impact on officials of stressful events may be a function of an official's ability or age. This observation states primarily that the younger the official, the more stressful officiating becomes (Anshel & Weinberg, 1995; Kaissidis & Anshel, 1993; Taylor et al., 1990). However, officials with more advanced levels of certification – as opposed to advanced age – reported higher levels of stress (Goldsmith & Williams, 1992; Taylor et al., 1990). Such conclusions show that interaction between these factors is, at best, tenuous, and that further research is needed in this area. Nonetheless, it would be reasonable to expect those officials with advanced levels of accreditation to be officiating in higher levels of competitive sports. This being the case, the greater emphasis on winning that is manifest in these levels, plus the likelihood of greater crowd numbers, increases further the pressure on referees to make correct decisions.

Although the actions of players and coaches have been implicitly, and explicitly, blamed for making a major contribution to officiating stress (Anshel & Weinberg, 1995; Taylor & Daniel, 1988), officials can unwittingly be the cause of player and coach aggravation, which can contribute to their own stress. Players and coaches are more likely to argue with an official who: lacks self confidence; makes inconsistent rulings; makes mistakes in positioning; and, shows a lack of concentration (Kaissidis & Anshel, 1993). Under such circumstances, officials may need to examine why hostility is being directed toward them before blaming players and coaches for being placed under undue pressure. Supporting this concept, Martens (cited in Anshel, 1989) stated, “referees on occasion may unintentionally foster aggressive behavior of participants, at least partly due to the nature of their interaction with players” (p. 32).

Despite the recognition of many stressful factors, situations, and events associated with officiating, in recent times some researchers have hypothesised that the relative magnitude of stress experienced by officials is not high (Anshel & Weinberg, 1995; Gait, Cook, Allen, & Duncan, 1979; Kaissidis & Anshel, 1993; Rainey, 1995; Rainey & Hardy, 1997; Rainey & Winterich, 1995; Stewart & Ellery, 1996). While these studies did not deny the existence of officiating stress, they questioned previously reported high stress levels that accompanied officiating work. Rainey and Hardy (1997) concluded, “there is growing support ... that most sports officials do not experience much stress associated with their duties” (p. 730). However, the same authors intimated that research into this area may be better served if the research designs were altered to assess officials' stress to specific game experiences, rather than the common practice of respondents providing a summative assessment of stress levels at the conclusion of a season.

Aside from the contradictions concerning the effects of stress, it is apparent that a number of factors contribute to stress experienced by officials. Moreover, some officials are affected to such an extent that the stress is manifested in illness and disease. Although the effects of such problems are experienced most profoundly by individual officials, there is an associated residual effect felt by sporting organisations as they attempt to meet increasing demands for qualified officials. Specifically, it is claimed that officials feel the pressure of balancing officiating responsibilities with private and professional lives (Goldsmith & Williams, 1992; Purdy & Snyder, 1985; Rainey, 1995; Zoller, 1985). Inevitably, people withdraw from the officiating role, and this has led to one of the most perennial problems confronting sport, i.e., the recruitment and retention of officials. Given the high demand for sports officials (Koskelainen, 1997), the effect of stress on officials should be of concern for all sporting organisations.

Recruitment and Retention of Officials

People take up officiating duties for many plausible reasons, including continued involvement in the game, financial rewards, challenge, excitement, power and commitment to fair play (Burke, 1991; Ferstle, 1978; Grunski, 1995; O'Bryant, 1996; Power, 1988; Purdy & Snyder, 1985; Spice & Thompson, 1994). However, the recruitment and retention of officials has been a problematic issue for many years (particularly in soccer, see Hay, 1999), and is still a major concern for numerous sports. These sports include soccer, hockey, baseball, rugby and Australian Football (Martin, 1996; Parkin, 1995; Power, 1997a; Spice & Thompson, 1995; Spice & Thompson, 1994; Taylor & Daniel, 1988; Webb, 1997). Exacerbating this persistent problem is the significant growth of competitive sports in recent years. Specifically, higher participation rates have created an "explosive demand for trained sports officials at every level" (O'Bryant, 1996, p. 32). O'Bryant's statement is illustrated clearly in the case of the Perth (Western Australia) metropolitan hockey competition, where during 1990, the weekend competition had only 55 independent umpires available to officiate 4850 registered players (Spice & Thompson, 1994).

However, the problem is not just recruiting officials, but also in retaining their involvement for the long-term. Parkin (1995) reported that Australian Football in the Victorian Country League expect to lose 50% of recruited officials in the first year. Similarly, 75% of young soccer referees (aged 15) have resigned by the time they are 18 years old (Power, 1988). Such statistics indicate the stark reality of the problem facing sporting organisations in supplying qualified officials for sporting competitions.

The reasons for low recruitment and retention rates are varied. Explanations offered include: no obvious career pathway; poor facilities at training and fixtured matches; lack of understanding of the role of the umpire by coaches and administrators; a lack of protection at matches; dissatisfaction with promotion systems; and, the negative attitudes toward officials

(Fry and Sefton, cited Parkin, 1995; Spice & Thompson, 1994; Taylor et al., 1990). Additionally, it is hypothesised that due to the stressful nature of officiating, novice officials who lack confidence may leave officiating before they have developed coping mechanisms required to deal with officiating stress (Taylor et al., 1990). In this regard, the effects of continual antagonism directed at officials is summarised by Knox (1998, p. 50) who stated:

In time, who will want to be a referee or umpire? Who will take the abuse, when the abuse comes from all corners? Nobody in their right mind would volunteer to be a public scapegoat.

These points highlight the magnitude of recruitment and retention problems currently facing sporting organisations. However, some sporting organisations have recognised the problem, and are now taking active steps to recruit and retain officials (Parkin, 1995; Power, 1997a; Webb, 1997). These include the creation of explicit and achievable career paths, the recruitment of officials through schools and tertiary institutions, the creation of a good career image, the provision of specialist officiating coaches, and adequate remuneration.

Of these strategies, appropriate financial reward is cited as a prime motivation to officiate (Power, 1988; Purdy & Snyder, 1985; Spice & Thompson, 1994). To illustrate this point, 23% of hockey players, surveyed by Spice and Thompson (1994), indicated that they would not take up officiating when their playing careers had finished, but may change their mind if remuneration was increased. Additionally, hockey umpires surveyed in the same study claimed that they were dissatisfied with the current pay structure. The remuneration issue was crystallised by Parkin (1995). He detailed how elite Australian Rules Football umpires received only a fraction of player salaries (approximately 10%). Given the important role officials play in the sport – particularly elite sport – such discrepancies are difficult to explain.

However, the value of increasing remuneration is debated by a number of sources. It was argued by Rokosz (1989) that higher payments do not attract necessarily a higher percentage of quality officials (although this source drew no conclusion related to the possible improved *quantity* of officials). Moreover, Power (1997a) maintained that cash rewards are only effective when the individual finds officiating enjoyable, and payments equal or exceed what could be earned elsewhere (e.g., casual work, overtime). Some perspective is placed on the debate by Spice and Thompson (1994) who concluded that extrinsic reasons may have some impact on players entering officiating, but intrinsic reasons are the major motivating forces for the continuation of officiating duties.

In summary, the conundrum of how to recruit and retain officials is vexed. It would seem that there is no easy or obvious answer for sporting organisations. Although higher salaries may go some way to alleviating the problem, there is no clear evidence that this strategy would lead to improved retention, and, more importantly, improved officiating standards.

What is clear is that sporting organisations need to examine the recruitment and retention problem from a number of perspectives, and, subsequently, develop a variety of strategies to maximise officiating personnel.

Summary

Although rule administration remains a core role of the official, numerous diverse issues now impact on modern officiating. The threat of litigation, allegations of bias, conflict with players and coaches, and performance assessment, are just some of these issues. Moreover, these issues affect officials across the sporting spectrum, irrespective of competitive level. As a consequence, many officials are not able to deal with the stress that such circumstances create, and leave officiating. This has given rise to a substantial vacuum of qualified and experienced officials, thus forcing sporting organisations to undertake active recruitment strategies to fill this void.

The necessity to prepare officials adequately to meet these challenges should begin with training and development programs. Programs should be comprehensive in their content, and tailored to the requirements of the sport and level of competition. Importantly, if training and development programs are to be contextual, the specific officiating competencies that are pertinent to a sport need to be identified. The next section examines this issue, with particular emphasis on specific competencies and their emphasis within individual sports.

THE PERFORMANCE CRITERIA OF A 'GOOD' OFFICIAL

This section analyses the performance criteria that officials require in order to demonstrate proficiency in their officiating role. Initially, discussion centres on the diverse competencies, skills and attributes perceived as important to officiating. This is designed to clarify the complexity of officiating work. Later discussion examines three aspects seen to be fundamental to officiating performance, namely: decision-making; fitness; and, communication. These performance requirements are explored independently, although their inter-relationship should not be overlooked, especially in dynamic team sports (e.g., soccer).

General Officiating Attributes

In the first section of this chapter, it was noted that there is little agreement about the performance requirements of officials (Clegg & Thompson, 1993). To some extent, this is a function of the lack of research (particularly compared to other participants in sport such as players and coaches) into officiating and related issues. As a consequence, when officiating performance is described, it is offered invariably in a global context, with no consistent or intimate understanding of which specific skills, competencies, and attributes officials should possess and display.

However, there have been some recent attempts to document performance criteria. These include personal qualities such as: rapport; decisiveness; courage; poise; integrity; judgment; confidence; effective self management; the application of personal health and safety principles; enjoyment; and, motivation. Added to this list are technical and mechanical officiating skills including: effective decision-making; effective planning and organisation; reaction time; hustle (move or act with resolute energy); and, the application of technical knowledge (Clegg & Thompson, 1993; National Officiating Program, 1997; Weinberg & Richardson, 1990).

Although the list of performance criteria is varied and seemingly comprehensive, it is not specific to any one sport or competitive level. The differentiation between generic skills and sport-specific skills is important. While there are underlying skills that all officials require, there are specific skills needed by officials to excel in their chosen sport. Consequently, officiating criteria that are sport-specific require identification and clarification.

Some attempts have been made in deference to this need. Ittenbach and Eller (1988), in an investigation of 58 American Football officials (Southern Conference), reported that officials distinguished the attributes of consistency, knowledge, ethics, self confidence and mental alertness as essential for good officiating. Subsequently, the 1990s saw additional attempts to identify specific competencies for specific sports (Anshel, 1995; Anshel & Webb, 1991). Specifically, Anshel and Webb (1991, pp. 35-36) identified eight performance categories which were seen as essential for Touch officials. These were:

1. Demonstrates mastery of rules.
2. Displays effective verbal communication skills on field.
3. Uses proper mechanics.
4. Displays effective communication skills off the field.
5. Demonstrates high level of fitness.
6. Actively contributes to the game.
7. Actively tries to improve own standards and skills.
8. Maintains credibility.

Although the research of Anshel and Webb (1991) identified sport-specific performance criteria, they did not examine criteria that related to the *competitive level* of the sport, i.e., there may be different skills required to officiate at the local junior level than at the senior representative level of a sport. The need to differentiate performance criteria in this manner was recognised by Soccer Australia, which undertook to clarify officiating competencies and skills needed at different competitive levels of soccer (see Appendix 1). Soccer Australia has nominated eight levels of refereeing competence, ranging from the beginner to international referee (Power, 1997a). Each level has expected performance criteria that referees must demonstrate before they move to the next officiating level. In this regard, the performance criteria are both sport and level specific. However, the criteria are not empirically derived, and, to some degree, lack specificity, rigour, and validity .

Notwithstanding these recent attempts to clarify sport specific officiating criteria, the officiating literature does focus upon three aspects of officiating performance which appear to be central to the officiating process. These performance criteria, i.e., communication, fitness, and decision-making, vary in their degree of application in different sports. However, the importance and inter-relationship which exists between these officiating skills is far-reaching, and is detailed in the following discussion.

Communication

The requirement for officials to possess proficient communication skills is well recognised in the literature (Anshel, 1995; Anshel & Webb, 1991; Evans, 1994; Weinberg & Richardson, 1990), and is viewed as a fundamental officiating skill (Australian Coaching Council, 1996). Weinberg and Richardson (1990) observed that “outstanding officials communicate effectively ... and clearly and consistently send the right messages” (p. 27). In doing so, the official conveys “confidence, control, calmness, positive intensity and fairness” (Weinberg & Richardson, 1990, p. 27). The importance of officials having well developed communication skills is further highlighted by the Australian Coaching Council (1996). It remarked that the official’s ability to relate to players, coaches administrators and other officials can enable the official “develop a positive and interactive social climate and foster the spirit of the game or event” (p. 6).

Officials use various methods of communication to convey their decisions and feelings. Most obviously, officials communicate through verbal interaction. However, it is not normally *what* is said that delivers the important message, but rather *how* the message is delivered (Evans, 1994). This method-of-delivery is referred to as “paralanguage” (Weinberg & Richardson, 1990, p. 32), and is elaborated upon by Anshel (1989). He maintained that a calm manner, in complex game situations, conveys a sense of control and maturity to both players and coaches. A positive application of paralanguage was encouraged by the Australian Coaching Council (1996) and Bies (1990), who both urged officials to speak to game participants in a friendly, yet firm, manner.

However, in relation to these latter recommendations, the effectiveness of assertive communication by officials should not be overstated. Anshel (1989) explained that an official who communicates in a “highly assertive, hardnose and insensitive manner may be displaying low, rather than high self confidence” (p. 33). Additionally, paralanguage used to patronise players and coaches does not reflect quality officiating, while sarcastic comments were seen as unprofessional, and have no place in sports officiating (Anshel, 1989).

Paralanguage application can be seen in non-verbal communication also. This can include hand signals, body language, spatial behaviour, whistle use and facial expressions

(Australian Coaching Council, 1996; Baer, 1990; Evans, 1994; Steel, 1993). Baer (1990) supported the notion of firm hand signals communicating conviction in a referee's decision, while Steel (1993) advocated the view that the official's facial expressions can convey strong messages to players. Similarly, Anshel (1989) suggested that certain gestures and mannerisms of the official help to reduce the inherent stress in the sporting environment. Overall, the evidence suggested that paralanguage, either verbal or non-verbal, presents officials with a powerful and flexible communication tool for conveying decisions.

Ostensibly, the bulk of communicative actions is to relay officiating decisions about game incidents. However, it is argued that communication needs to be more extensive than the appropriation of fair or foul play. Officials should also be prepared to articulate a rationale for their decisions. Askins et al. (1981) contended that "officials must communicate ... to the players and coaches that their [the officials'] connection between the incident and the rule is appropriate" (p. 91). This stance was endorsed by the Australian Coaching Council (1996), which listed this function as one of its communication "hints for officials" (p. 10). However, the complexity of this form of communication requires an understanding deeper than the generation of simple hand signals. Particularly, it obligates officials to develop communication skills that promote reason and consideration. This is not an easy task. Nonetheless, the willingness of officials to explain their decisions can lead to associated side-benefits. Anshel (1989) noted that referees who undertake such communication strategies were more likely to gain a greater appreciation of their work, particularly from players and coaches.

Although discussion thus far has centred on an official's communication with game participants, it is recognised that the nature of communication – between officials and game participants – should not all be one-way. Communication needs to be a two-way process, with the official willing to listen to officiating concerns expressed by game participants. As noted in the previous section, the type of interaction that ordinarily occurs between officials and game participants is somewhat hostile and is a prime source of officiating stress. However, it is argued that officials "would markedly reduce coach and player hostility if they indicated ... an interest in their [players and coaches] comments" (Anshel, 1989, p. 33). The need for effective two-way communication is summed up by Weinberg and Richardson (1990, pp. 28-29) who stated that officials should:

Listen with an open mind! You [the official] will not communicate effectively if you are evaluating or judging the message as it is being delivered. A common mistake is to stop listening.

In light of the scenarios and circumstances described above, the ability of the official to communicate through effective and meaningful methods is paramount for successful officiating. While superficial communication skills are necessary for relaying technical decisions, the more complicated skills that impinge on player-management are less easy to

develop. Moreover, advanced communication skills allow the official to gather more evidence by which to make accurate and fair decisions (National Officiating Program, 1998).

Fitness

Officiating sport, particularly those sports that are fast moving and/or undertaken on expansive playing dimensions, requires high levels of aerobic fitness. This requirement is even more pertinent when the official is required to maintain close proximity to the ball or play. Heart rate and maximal oxygen up-take figures, emanating from research across a number of sports, supports this need. According to Conti and McClintock (1983), American Football referees achieved heart rates of 77.9-98.9% HR_{max} , with mean heart rates of head referees ranging from 141 $b \cdot min^{-1}$ to 161 $b \cdot min^{-1}$. Netball umpires spend almost half of the game with a heart rate of 75-85% HR_{max} (Otago et al., 1994), while mean heart rates of 159 $b \cdot min^{-1}$ have been reported for Australian Football umpires (Pyne & Ackermann, 1987).

Similarly, high aerobic fitness requirements are demanded for soccer referees. An examination of the distances covered in a game by soccer referees, and the movement patterns which constitute these distances, underscores the need for high levels of fitness. Movement patterns, such as walking, forward running (jogging and sprinting), reverse running, and sideways movements (Asami, Togari, & Ohashi, 1988; Catterall, Reilly, Atkinson, & Coldwells, 1993; Ferstle, 1978; Friedman & Klein, 1988; Johnston & McNaughton, 1994; Reilly, 1996) can result in the referee covering up to 11 km per game (Reilly, 1996). A summary of the varying distances, and associated movement patterns are outlined in Table 1.3. Movement patterns and distances, similar to those listed in Table 1.3, have also been reported for officials in netball, Australian Football, and Touch (for example, see Craig, Jaques, Pavia, & Squires, 1979; O'Connor, 1994; Otago et al., 1994; Pyne & Ackermann, 1987; Smith, 1980).

It is of note from Table 1.3 that the majority of distance covered by referees involves a jogging movement pattern. Therefore, it can be extrapolated that the major physiological demand placed on a referee is by the oxygen transport system (Asami et al., 1988; Catterall et al., 1993; Reilly, 1996). Although these authors did recognise that there should be some emphasis given to the anaerobic energy system, the strong demands placed on the oxygen transport system are supported by unequivocal empirical research. Johnston and McNaughton (1994) reported mean heart rates for Tasmanian State League soccer referees to be 163 beats per minute ($b \cdot min^{-1}$) for the first half and 162 $b \cdot min^{-1}$ for the second half, and mean heart rates (over the full match) of 165 $b \cdot min^{-1}$ were reported for Class 1 soccer referees in Britain (Catterall et al., 1993). Such heart rate figures show that soccer referees

Table 1.3 Movement patterns and distances covered* by soccer officials
(Percentages in brackets where provided by author)

Author	Distance Covered (mean)	Walking	Jogging	Fast Running	Reverse Running
Asami et al. (1988)	9990	3385 (33.8)	4851 (48.5)	1772 (17.7)	1081 ¹ (10.8)
Friedman and Klein (1988)	-	5000 ²	-	400	700 ³
Catterall et al. (1993)	9438	2613 (23.0)	4444 (47.0)	1109 (11.8)	1722 (18.2)
Johnston and McNaughton (1994)	9408	1780 (18.9)	4382 (46.6)	1725 (18.3)	1521 (16.2)
Reilly (1996)	11200	-	-	-	-

* meterage per game

¹ Stated distances for walking, jogging and sprinting.

² Combined figure represents walking and running at different speeds.

³ Combined distance with sideways movement.

are undertaking work that is at least 70-85% of maximal heart rate (HR_{max}), and, in many cases, the majority of the match is spent with a heart rate greater than 85% HR_{max} (Johnston & McNaughton, 1994). For elite World Cup soccer referees, the world governing body of soccer, the Federation Internationale de Football Association – FIFA (1995) – reported maximal heart rates as high as 200 $b \cdot min^{-1}$. Given the added psychological demands of refereeing, and that referees are, on average, 25 years older than players, the physiological requirements of refereeing are significant.

A second perspective on referee fitness was provided by Ferstle (1978). He claimed that soccer referees tended to do more running than players. Although surprising at face value, this observation was supported by FIFA (1995), who, more specifically, acknowledged that referees undertake a greater amount of sprinting and fast jogging than midfield players. It was hypothesised that this circumstance arose because referees constantly follow play, while players have some degree of choice in undertaking high-intensity effort 'off-the-ball' (Reilly, 1996). As a result, players can gain rest periods from time-to-time during the course of a game, whereas referees do not have such a provision due to the nature of their role. Johnston and McNaughton (1994, p. 72) concluded that:

Referees should be treated as legitimate athletes and should undertake their training accordingly and need intensive training regimes, with the emphasis on aerobic exercise, if they are to meet the demands imposed in a typical soccer match.

Although the importance of good physical conditioning can not be understated, a side-comment provided by Parkin (1995) is also pertinent. He gave prominence to the need for officials not only to be fit, but to look fit too. Parkin stated the importance for officials to

display a “first class physical image,” and that the image of the official suffers if they “present an image of obesity and minimal physical fitness” (p. 20).

Overall, the physically demanding nature of most forms of officiating means that officials need to give close attention to their physical conditioning. Importantly, the official should undertake physical training which reflects the physical demands of the game (Johnston & McNaughton, 1994; Soccer Australia, 1995). More often than not, this should focus predominantly on the aerobic energy system. By maintaining appropriate fitness levels, officials are able to sustain respect from players, coaches and spectators, while making decisions that are made from the best possible position.

Decision-making

Decision-making is central to the role of officiating (Evans, 1994). This was illustrated by Van Meerbeek, Van Gool and Bollens (1988), who, in an analysis of refereeing decisions during the 1986 Soccer World Cup, reported that referees averaged one decision every forty seconds. Additionally, in a separate report, it has been estimated that 90% of decisions that an official makes are ‘judgement calls’, i.e., decisions which are at the discretion of the official by (Ferstle, 1978). Within this decision-making context, it was asserted that decision-making should always be firm, clear, impartial and decisive (Bies, 1990; Reilly, 1996; Wilson, 1986), as decisiveness can turn controversial or discretionary decisions into accepted decisions (Clegg & Thompson, 1993). These points and examples illustrate the fundamental importance of decision-making to effective officiating

The often subjective nature of officiating makes the interpretation of most game rules and laws open to debate. Rains (1984) noted that referees are given flexibility to make decisions, and that many decisions are made in the context of a game. The game context is constituted by a number of factors, including: the game’s immediate history, i.e., what has happened previously (including decisions that have already been made concerning rule infractions and any subsequent sanctions); the game’s unfolding future; and, the consequences that follow any decisions that are made from that point in time. All these factors impact on officials in different ways, and, consequently, can result in different decisions for an apparently identical incident. However, despite the contextual nature of decision-making, it is acknowledged that consistent decision-making is a hallmark of good officiating (Ittenbach & Eller, 1988; Rains, 1984).

The contextual nature of decision-making is illustrated clearly in a sport such as soccer. The Laws of Soccer (Soccer Australia, 1998) are universally agreed upon by national soccer organisations, and are administered by FIFA. However, as pointed out by Wilson (1986), any uniformity in the interpretation of the Laws is difficult to achieve. Primarily, this is because no two decisions are exactly alike, and other factors (such as climatic conditions, the

type of ground, the referee's personality, the skill of the players, the quality of the match, the conduct of players, club officials and spectators) always represent non-constant variables (Wilson, 1986). Similar variations in rule interpretations were observed in other sports also. Specifically, Australian international softball umpires reported that the interpretation of rules within Australia is considerably different in practice from international rules (Koskelainen, 1997).

In addition to the reasons for variability in officiating decisions offered by Wilson (1986), it was maintained by Murphy, Tanzor and Trotter (1995) that officiating decisions are dependent on two constructs, namely, memory (past experience) and strategies (undertakings that have, or have not, worked). The effect of these constructs are further compounded by other factors, including: attention; individuality; practice; motivation; fitness; and, environment (Murphy et al., 1995). Given these constructs and contributing factors, the expectation of players, coaches and spectators that officiating decisions should be highly accurate and consistent under all circumstances is unrealistic (Murphy et al., 1995). Yet, these issues aside, it is asserted that sports officials are required to know the technical aspects of laws and rules, *and* to be able to apply them in ambiguous game situations (Ittenbach & Eller, 1988).

From a broader perspective, the symbiotic relationship between decision-making and other officiating performance criteria in soccer is noteworthy. Fitness, in particular, has been closely linked to effective decision-making. Asami, Togari and Ohashi (1988), observed that soccer referees are required to cover a wide range of the playing area, while simultaneously making good decisions. By implication, decision-making can be prone to error where high levels of physical intensity are experienced (Reilly, 1996). As concluded by Soccer Australia (1996), the fitter the referee, the quicker and more accurate are the decisions given. It is therefore not surprising that physical fitness was identified as one of the top three officiating priorities by international soccer referees (Friedman & Klein, 1988).

To clarify this concept further, the proximity of the official to game incidents is crucial for making the correct decision. Without doubt, it is fundamental for officials to be close enough to play to obtain all relevant information concerning game incidents. Without full possession of the facts, decisions are based, to some degree, on circumstance and conjecture. This not only increases the likelihood of the decision being incorrect, it can diminish the value of the official's performance. The relationship between fitness and decision-making is probably best encapsulated by Brown (1993, P. 17) in her commentary on netball umpires:

Coaches and players appreciate umpires who are fit enough to keep up with play and position well to judge ... critical incidents. They [players and coaches] deplore the lazy umpire who guesses.

The linkages between decision-making and fitness with communication are less obvious, but, nonetheless, important to effective officiating. In this regard, the fitness of the official plays a pivotal role. Firstly, the official is more likely to communicate effectively when they are confident in their decisions. This is often a result of the official obtaining good positions to view game incidents, which, in turn, is a function of the official being fit enough to 'keep-up' with play. Secondly, the articulation of decisions, and the willingness of the official to engage players and coaches in discussion, are more emphatic if the official is not breathless from physical exertion.

In conclusion, the numerous factors which contribute to officiating decisions have an obvious impact on the perceived ability and integrity of the official. If it is accepted that the primary role of the official is to uphold and apply the game rules, any major variance in rule application and administration impact on the credibility of the referee. As discussed in the previous section, issues relating to officiating bias, player/coach relationships, and litigation all have their foundation in the decisions that officials make. Moreover, the decisions they make directly impact on performance assessment. This being the case, the relationship between the decision-making ability of the official, to his or her effectiveness, is clear.

Summary

The qualities needed by modern sports officials are diverse, and subsequently, not easily categorised. Aside from communication, fitness, and decision-making skills, and the unavoidable interaction of these performance criteria, the criteria required for officiating seem inexhaustible. The reasons for this diversity include the nature of the sport, the associated competitive level, the social context in which the sport is being played, and the expectations of those participating in the sport. If it is accepted that these parameters impact on the performance criteria needed by officials to undertake their role in a proficient manner, then it is perhaps somewhat impractical to expect that a definitive list of officiating criteria, which encompasses all officials at all levels of all sports, can be developed.

Seemingly, it would be more beneficial to examine the nature of individual sports within their competitive contexts. This would allow a more valid identification of relevant and appropriate skills. As such, officials would have a better idea of what is expected of their performance, while players, coaches, spectators and other interested parties will have potentially a deeper understanding of the officials' role and responsibilities.

CONCLUSION

In recent years there has been an improved recognition and acceptance of issues confronting the officiating profession. As sport has become a more integrated part of our society, the sporting spotlight has been focused on the official more often than in previous times. Closer analysis of the responsibilities of the official, and the subsequent impact of officiating

decisions, has facilitated this change. No longer is the official's decision seen as 'final,' nor is the integrity of the official beyond reproach. Officiating performance is now under constant scrutiny from athletes, coaches, spectators and the media.

Such scrutiny has led to a number of challenges for officials. It has become apparent that officials now need to be able to deal with allegations of officiating bias, player and coach misconduct, litigation, and the subsequent stress that is derived from meeting these challenges. Moreover, the serious problem of recruitment and retention of officials may be symptomatic of officials unable to cope with the peripheral side-issues now beginning to permeate modern officiating.

A comprehensive understanding – by all sporting stakeholders, including players, coaches, administrators and officials – of the roles, competencies, skills, and responsibilities required by officials would help officials meet these challenges. However, it is not entirely clear just what these performance criteria are. Essentially, the *research* literature relating to the basic characteristics of effective officials is not extensive. The review of literature has identified that there are significant gaps in the research related to officiating. Reasons for this lack of information are not clear, although one can hypothesise that the role of the official has, traditionally, not been seen as important as the athlete or coach.

Given these circumstances, there is no concise overview of what constitutes a 'good' referee or umpire, or how such officials are trained and developed. This is despite such understandings being pivotal to the improvement of officiating standards, and the validation of the officials' role. Clearly, there are broad performance areas which are common to most officials in most sports (National Officiating Program, 1997). However, the estimation and development of specific performance criteria for specific sports is more problematical.

Similarly, recognition that officials may require different skills (and undertake different roles) at various levels of officiating is not apparent. The scant literature that is available verifies this point. Research has tended to focus on a relatively small number of sports (for example, see Anshel, 1995; Anshel & Webb, 1991), and does not focus on one specific competitive level. As a consequence, it is difficult to apply results and conclusions to specific levels of officiating. Moreover, it is spurious to infer which performance criteria are applicable to other sports. There is perhaps an implicit understanding that some criteria, such as communication and decision-making, are not sport-specific. However, empirical research substantiating this concept is not evident.

Ostensibly, the performance criteria that officials are expected to display, at *any* level of competition, cannot be expected to be intrinsically developed. It is clear from the literature that knowing the rules of the game is no longer a sufficient prerequisite to officiate. Therefore, any training or ongoing education programs need to incorporate the development

of all performance requirements. However, the literature does not provide comprehensive models for such development, nor does it detail the theoretical underpinnings of officiating training and development. This lack of detail is an impediment to a more professional approach to officiating, and the desire of officials to be accepted as a legitimate and vital component of sporting contests.

Officiating training and development is not the only aspect of officiating that is being restrained by a lack of specified performance criteria. Problems with assessment are also evident. Officials are normally evaluated by a number of sources, the most common being officiating inspectors and coaches. However, the criteria that assessors use are neither consistent nor based on a sound research perspective. The development of explicit performance criteria has the potential to reduce this problem. This would remove, to some degree, subjectivity from the assessment process. Moreover, it would allow officials to have a clear understanding of how they are being rated, and provide assessors with meaningful and valid assessment criteria.

Clearly, advancement in the standards of any occupation is predicated on an unambiguous and consensual understanding of performance criteria. From this perspective, officiating is no different to any other occupation. Particularly, the identification and development of specific performance criteria represent a foundation for effective officiating. The potential benefits this brings to related aspects of officiating cannot be ignored. However, the problem of how to identify essential criteria remains. The following chapter details a technique – the Behaviourally Anchored Rating Scale – that can be used to address this problem, and, as such, can be used as a starting point to profile the work of officials.

CHAPTER TWO

IDENTIFICATION OF OCCUPATIONAL PERFORMANCE CRITERIA

Introduction

A recurring theme in the literature has been the need to delineate and identify explicitly performance requirements for effective officiating. This was confirmed by Anshel (1995, p. 9) who stated, “relatively few attempts have been made to systematically quantify the criteria by which researchers and practitioners interpret and assess effective or desirable performance in sport.” For determining relevant officiating performance criteria, for any sport at a given competitive level, Anshel’s quote is brought into sharper focus due to the dearth in related research. Moreover, the problem is compounded by the evolution of officiating into an increasingly complex profession, thus necessitating a deeper analysis and understanding of officiating work.

Despite the lack of empirically derived officiating standards and expectations, there is no shortage of informal opinion concerning the quality of officiating performance. The existence of vague and imprecise ideas on what constitutes effective officiating performance is apparent following the conclusion of any sporting event, but the foundation of such opinion is based rarely on definitive criteria. Therefore, the issue especially relevant to officiating is the identification of specific and relevant performance criteria that are both valid and reliable representations of officials’ work.

One method which addresses such an issue is the Behaviourally Anchored Rating Scale (BARS). The primary purpose of BARS is to determine “valid and reliable job criteria to assess job performance as derived from behavioral evidence of effective performance” (Anshel, Housner, & Cyrs, 1987, p. 81). In light of this definition, the aim of this chapter is to examine BARS procedures for the development of essential occupational performance criteria.

The chapter is divided into two sections, and explores BARS methodology from two perspectives. In the first section, the background of BARS is reviewed, with particular reference to benefits derived from the methodology. The second section investigates the development and applications of a hybrid BARS variant, and how this newer approach articulates with traditional BARS procedures.

BACKGROUND TO THE BEHAVIOURALLY ANCHORED RATING SCALE

The Behaviourally Anchored Rating Scale (BARS) was developed originally by Smith and Kendall (1963) to evaluate the performance of head nurses. Since its inception, it has been utilised in the development of performance criteria across many diverse roles. Examples include faculty teaching (Bernardin, 1978; Harrai & Zedeck, 1973; Hom, DeNisi, Kinicki, & Bannister, 1982), policing (Bradley & Pursley, 1987), firefighting (Dickinson & Tice, 1973), seamanship (Borman & Dunnette, 1975), university teaching (Conway, 1995; Divoky & Rothermel, 1988; Harrai & Zedeck, 1973; Hom et al., 1982), and primary school teaching (Moore & Webb, 1995). Furthermore, it has been used to identify and measure psychological parameters such as stress (Shirom, 1988), time urgency (Landy, Rastegary, Thayer, & Colvin, 1991) and feedback quality (Hom et al., 1982). However, despite the diversity of BARS utilisation, its application to sporting research is not common. Sporting applications have been relatively recent, and included one study investigating coaching competencies (Anshel et al., 1987), and two investigating officiating competence (Anshel, 1995; Anshel & Webb, 1991).

As the name implies, BARS is essentially a rating scale instrument. However, it differs from many other rating formats in that performance is captured in “multidimensional, behavior specific terms” (Schwab, Heneman, & DeCotiis, 1975, p. 550). Importantly, it provides narrative descriptions of the constituent of excellent quality work (Leap & Crino, 1993). Implicit in these characteristics, and central to the effectiveness of BARS, is the methodology’s ability to identify specific occupation-related performance dimensions, plus provide unambiguous examples of these dimensions (Anshel et al., 1987; Jacobs, Kafry, & Zedeck, 1980; Landy et al., 1991; Leap & Crino, 1993; Schwab et al., 1975). Similarly, a focus on multidimensional occupational *performance*, as opposed to occupational *outcomes* (Stoskopf, Glik, Baker, Ciesla, & Cover, 1992), complements the analytic foundations of BARS.

These attributes of BARS have their roots in the critical incident technique developed by Flanagan (1949). This comprised the creation of checklists which contained occupational behaviours. The behaviours were viewed by supervisors and subordinates as both essential for exemplary performance and indicative of unsatisfactory performance (Jacobs et al., 1980). However, BARS methodology took Flanagan’s procedure an additional step by classifying behaviours into occupational performance dimensions. This allowed performance to be viewed from a broad perspective, while still providing detailed examples of specific work requirements.

Fundamental to this additional step are a number of procedures designed to preserve the integrity of resultant performance measures. The following discussion provides an overview

of these procedures, with analysis specifically focused on how the technique of Flanagan (1949) was extended to develop performance dimensions. In doing so, discussion draws heavily on the original BARS of Smith and Kendall (1963).

Development of Behavioural Anchored Rating Scales

The scales developed first by Smith and Kendall (1963) were termed Behavioural Expectation Scales (BES). This was because work descriptions were written in as *expected* behaviours, rather than *observed* behaviours. However, such an expression of behaviours was altered subsequently by Dickinson and Tice (1973) to provide more definitive examples of work behaviour, thus the generic 'BARS' term was popularised and became generally accepted.

Notwithstanding semantic name changes, BARS development occurred through an iterative process, and typically involved five predetermined steps. These steps, outlined below, are based on the original descriptions by Smith and Kendall (1963). The completeness and order of these steps have remained relatively unchanged in subsequent years, however, some minor alterations have become evident, and are noted accordingly (for example, see Campbell, Dunnette, Avery, & Hellarick, 1973; Schwab et al., 1975).

1. This initial step involved the development of *performance dimensions*. Dimensions were defined as general qualities or characteristics of the role under investigation (Smith & Kendall, 1963). The task of generating dimensions was undertaken by groups of people who were familiar, knowledgeable and/or possessed expertise in the profession or role under investigation (i.e., profession experts).
2. Following the identification of performance dimensions, *critical incidents* were then identified. This task required the profession experts to describe specific behaviours that illustrated both effective and ineffective performance within each dimension.

These two steps were altered subsequently in some later studies (e.g., Bradley & Pursley, 1987; Campbell et al., 1973). This alteration was seen as a means of keeping participants focused on specific behaviours, i.e., critical incidents, rather than on traits or global performance dimensions (Campbell et al., 1973).

3. The third step is referred to as the *retranslation* process (Smith & Kendall, 1963). This stage required the re-assignment of incidents back into performance dimensions, a task undertaken by a second group of relevant profession experts. This group was given the list of critical incidents developed originally in Step 2, and asked to place incidents back into the dimensions in which they were classified originally (Landy et al., 1991). Normally, an incident is retained if 50-80% of the group (criteria varies

- between studies) reassign the incident into the same dimension as did the first expert group (Schwab et al., 1975).
4. Following retranslation, incidents were *scaled*. Specifically, incidents were assessed to determine their relevance to performance. The process was undertaken by the second group of experts, who denoted (usually on a 7-or-9 point scale) how effectively, or ineffectively, incidents described performance. The mean values that resulted from this process indicated where the incident was positioned on the scale. Incidents were retained if they were contained within a predetermined standard deviation. Once again, figures vary between studies, but standard deviations of 1.5 or less on a 7-point scale were generally accepted (Schwab et al., 1975).
 5. In this final stage, incidents remaining from retranslation and scaling were used as *anchors* for respective performance dimensions. Incidents were rewritten in the form of behavioural 'descriptors' or 'indicators,' and arranged in a vertical formation within their respective dimensions. The location of an indicator along the vertical scale was in accordance with the rating established for the incident in Step 4 (see Figure 2.1 for an example of this format). When used in a practical setting, raters selected an indicator that best resembled the performance of the ratee for each dimension (Grussing, Silzer, & Cyrs, 1979). This indicator, which corresponded with a number on a vertical scale, became the ratee's score for that dimension (Grussing et al., 1979).

These five procedures are viewed as a primary strength of BARS methodology. Through these processes, it was asserted that the final product provided a multidimensional job analysis (Blood, 1974; Dickinson & Tice, 1973; Schwab et al., 1975), where dimensions were designated consistently as important (Dickinson & Tice, 1973), and defined using unambiguous real-world performance examples (Grussing et al., 1979). Particularly, the use of profession-related experts was integral to the process. Their relevance was based on the premise that people who were cognisant with the area of performance under investigation brought a variety of perspectives to the process.

Accordingly, BARS has the potential to foster at least three prospective outcomes. Firstly, the language and grammar used in specifying performance criteria are relevant to the role under investigation. Implicit in this are performance dimensions and indicators which are written in the language of people who use them (Campbell et al., 1973). Items tend to tap employee slang and culture (Anshel et al., 1987), and are highly job-specific (Schwab et al., 1975). Therefore, performance descriptors are devoid of psychological jargon, the purpose being to make all items easily understood and meaningful to those who use the scales (Landy, 1985; Shirom, 1988). Likewise, the resulting terminology minimises possible ambiguity associated with the meanings ascribed to dimensions and indicators (Anshel, 1995; Borman, 1975; Campbell et al., 1973; Schwab et al., 1975).

	9—	—	Could be expected to conduct a full day's sales clinic with two new sales personnel and thereby develop them into top sales people in the department.
Could be expected to give sales personnel confidence and a strong sense of responsibility by delegating many important jobs to them.	—	—8	
	7—	—	Could be expected <i>never</i> to fail to conduct training meetings with employees weekly at a scheduled hour and to convey to them exactly what is expected.
Could be expected to exhibit courtesy and respect toward sales personnel.	—	—6	
	5—	—	Could be expected to remind sales personnel to wait on customers instead of conversing with each other.
Could be expected to be rather critical of store standards in front of employees, thereby risking developing poor employee attitudes.	—	—4	
	3—	—	Could be expected to tell an individual to come in anyway even though she/he called in to say she/he was ill.
Could be expected to go back on a promise to an individual who had told could transfer back into previous department if she/he didn't like the new one.	—	—2	
	1—	—	Could be expected to make promises to an individual about his/her salary being based on department sales even when the supervisor knew such a practice was against company policy.

Figure 2.1: Example of BARS performance dimension to assess the effectiveness with which managers supervise sales personnel

(adapted from Campbell et al., 1973)

Secondly, the use of occupational representatives in scale development was seen to increase the likelihood, and willingness, of individuals to cooperate with the use of the scales (Schwab et al., 1975; Shirom, 1988). As Schwab et al. (1975) noted, improved motivation resulted when individuals knew that they, or their representatives, participated in scale development.

Lastly, the BARS process forced participants to take a “long, hard look at performance” (Landy, 1985, p. 186). In doing so, stereotypical behaviours of poor performance or the unsuccessful worker are avoided (Landy, 1985). Only those actions and tasks that *do* occur in the successful day-to-day execution of an occupational role are listed as indicative of performance.

It would seem that BARS, through its rigorous developmental procedures, provided a useful and highly practical means of identifying relevant occupational performance requirements. Accordingly, the technique presents a *prima facie* case for its utilisation in performance assessment (the intended original use of BARS, see Campbell et al., 1973; Landy & Guion, 1970; Zedeck & Baker, 1972). However, as with any other assessment instrument, the success of BARS for this purpose required close scrutiny in terms of relevant psychometric criteria.

Psychometric Characteristics of BARS

Perhaps the most discussed and researched aspect of BARS was the methodology’s claimed positive effects on psychometric properties such as validity, reliability, halo error, and leniency error. Such criteria can pose a threat to the effectiveness of any rating instrument, and needed to be assessed adequately before BARS instruments could be applied confidently in practical settings. The following discussion examines BARS against these criteria, and draws on the extensive research literature (predominantly published during the 1970s) relating to these parameters.

Validity

The over-riding principle of validity is to indicate the degree to which an instrument measures what it intends to measure (Oppenheim, 1992). In this regard, validity concerns of BARS revolve around two forms, namely, construct validity and face validity. Construct validity demonstrates how well an instrument relates to a set of theoretical assumptions about an abstract construct, e.g., conservatism (Oppenheim, 1992). In BARS literature, construct validity is used interchangeably with the term discriminant validity, i.e., the ability of relevant performance dimensions to discriminate accurately between performance criteria. This confuses the interpretation of results somewhat, however, the general thrust of reported findings showed scale independence to be problematic.

Attempts to demonstrate empirically the conceptual independence of scales, predominantly through factor analysis, have realised limited success (studies presenting clear exceptions include Landy & Guion, 1970; Landy et al., 1991). Problems associated with extracting consistent non-trivial factors (Kafry, Jacobs, & Zedeck, 1979), low percentages of accountable variance (Dickinson & Trice, 1977), and high correlations between extracted factors (Keaveny & McGann, 1975) illustrated complications concerning scale independence. This has led to some studies questioning the ability of BARS to provide adequate discrimination between scales (Bernardin, 1977; Campbell et al., 1973).

These problems notwithstanding, one clear exception has been the work of Stoskopf et al. (1992). Using relatively rigorous decision rules, they derived a three-factor solution from 27 behavioural items. These three factors accounted for 84% of explained variance, with each factor containing at least six significantly loading items. They concluded that “because the factor analysis used identified three strong factors, and because the factors were conceptually meaningful, it is reasonable to suggest that the BARS ... has construct validity” (p. 341).

Such endorsement of BARS with respect to construct validity is not common. Nevertheless, it is asserted that “though development of BARS insures *conceptually* distinct dimensions, in actuality, performances among dimensions can be correlated” (Jacobs et al., 1980, p. 620). Moreover, it is seen as almost impossible to escape the pervasive influences of one dimension to the next (Burnaska & Hollmann, 1974), and, as such, total scale interdependence is perhaps an unrealistic expectation (Schwab et al., 1975).

The second form of validity under discussion, face validity, is premised on an instrument *looking* like it is measuring what it claims to measure (Crowl, 1993), e.g., an instrument designed to assess a child’s mathematical ability should appear to be measuring this mathematical ability. On this issue, there seems to be greater consensus for BARS to satisfy this parameter. The previous discussion on the use of profession-related experts in the development of scales showed clear evidence in this regard (Anshel, 1995; Landy, 1985; Smith & Kendall, 1963), i.e., this procedure ensures that all occupational criteria are covered, and that the terminology is in the language of the worker. Moreover, the *retranslation* phase of scale development is seen to further enhance face validity (Kavanagh & Duffy, 1978), as only items that satisfy specific agreement criteria are included in the scales.

In summary, the BARS technique has not been conclusive in meeting an important psychometric criteria. Although general acceptance exists for the establishment of face validity, definitive judgements concerning construct validity are problematic. Perhaps a solution to this latter concern exists with operational definitions utilised in the literature which define discriminate and construct validity. In brief, consistent approaches to the measurement of scale validity are required before definitive judgements on BARS validity can be made.

Reliability

In its most general form, reliability refers to “the ... consistency of a measure, to repeatability, to the probability of obtaining the same results again if the measure were to be duplicated” (Oppenheim, 1992, p. 144). When BARS methodology was first developed, it was claimed that scales provided exceptionally high reliability. Specifically, in the study by Smith and Kendall (1963), all scales had reliability coefficients of 0.98 or higher. However, these reliability findings from the research of Smith and Kendall, and from those that soon followed (for example, see Landy & Guion, 1970), were found to be somewhat problematic. In fact, later judgements of the scale reliabilities quoted by both Landy and Guion (1970) and Smith and Kendall (1963) were damning. Particularly, Zedeck and Baker (1972) described the findings as “artifactual and spuriously high” (p. 458).

Reasons for this criticism from Zedeck and Baker (1972) were attributed to specific procedures associated with the development of BARS instruments. They maintained that by rejecting items with large standard deviations (after originally selecting a wide range of items per dimension), it is not surprising that near perfect reliability scores were obtained. Borman and Vallon (1974) cited other procedures which could artificially inflate reliability results. These include: choosing anchors whose means are relatively far apart; having additional raters assign values to the anchors during the retranslation process; and, including a high number of anchors per dimension. Additionally, they hypothesised that high scale reliability co-efficients were a result of reliability being established during the *developmental* process of BARS, rather than scale reliability being assessed as a consequence of the *utilisation* of BARS in applied settings (Borman & Vallon, 1974).

This criticism aside, it is argued that debate concerning the establishment of reliability from either developmental or utilisation processes may be superfluous. It is argued that, from an applied perspective, inter-rater reliability is more important than scale reliability *per se* (Schwab et al., 1975). However, inter-rater reliability results for BARS have not been exceptional. Reported inter-rater reliability coefficients have shown wide variation in a number of studies, ranging from 0.17 to 0.74 (for example, see Bernardin, Alvares, & Cranny, 1976; Borman, 1975; Borman & Vallon, 1974; Zedeck & Baker, 1972). The ambiguous results from these early studies were encapsulated concisely by Jacobs et al. (1980, p. 617), who concluded:

From the compilation of inter-rater reliability for BARS it seems reasonable to conclude that the methodology results in varying modest levels of agreement between raters. This calls into question the ... expectation that BARS will increase reliability of assessments over other evaluation techniques.

Reasons for indifferent inter-rater reliability results are varied, but the most obvious is that cited by Schwab et al. (1975). They claimed that inter-rater reliability concerns may be

caused by critical incidents being discarded during *retranslation*. As this process resulted in numerous incidents being removed, some instruments may contain dimensions that are defined by relatively few behavioural examples. Consequently, confusion in the mind of the rater may be created due to observed behaviours not being listed in the scale (Schwab et al., 1975).

In an attempt to circumvent this problem, the training of raters in the use of BARS was hypothesised as a strategy for improving inter-rater reliability (Bernardin, 1978; Borman, 1979; Borman, 1975; Burnaska & Hollmann, 1974). However, it seemed that this strategy had little or no effect in improving inter-rater reliability. For example, Borman (1975) found that, as a consequence of rater training, halo error was reduced. But in achieving this end, the greater distinction between a single ratee's levels of dimension performance led to decreases in inter-rater reliability.

Later studies confirmed the contradictory trend. Bradley and Pursley (1987) reported inter-rater reliability coefficients ranging from 0.21 to 0.89 across eight performance dimensions for patrol officer performance, while Stoskopf et al. (1992) recorded co-efficients of 0.34 to 0.96 across 27 behavioural descriptors for the assessment of nurse assistants (interestingly, this latter study did not use performance categories, but instead used behavioural descriptors). However, findings reported by Anshel (1995) reversed this tendency toward inconclusive inter-reliability results. He used a hybrid version of BARS to measure officiating competence in basketball, and reported inter-rater reliability co-efficients of 0.79 to 0.88 across 13 performance dimensions.

It is apparent that the reliability of BARS instruments is open to debate. The exceptional scale reliability coefficients reported in early BARS studies have not been replicated. Moreover, the capacity of BARS to foster acceptable levels inter-rater reliability has been questioned. Given the contradictions which exist in the BARS literature concerning reliability, there is little reason to believe that BARS, in the form proposed by Smith and Kendall (1963), can provide consistently superior reliability results.

Halo Effect and Leniency Error

Although validity and reliability represent two important psychometric properties, other measures are equally important. Specifically, these are halo error and leniency error. Halo error refers to an individual being rated on the basis of a favourable or unfavourable general impression. Ratings are consequently assigned that are consistent with the impression, irrespective of merit (Landy, 1985). Leniency errors occur when an individual is rated higher, or lower, than is deserved (Borg & Gall, 1989). This results in a markedly skewed, either positive or negative, distribution of ratings (Jacobs et al., 1980). Halo effect and

leniency errors pose threats to rating accuracy, and their minimisation is central to assessing the integrity of BARS (Jacobs et al., 1980).

In examining the capacity of BARS to reduce halo error, results have not been conclusive. This is probably because, in part, research tended to make comparative assessments against other rating techniques. When compared to summated scales, contradictory findings emerged. Campbell et al. (1973) and Borman and Dunnette (1975) claimed BARS was superior, while other studies found no significant differences between the two techniques (e.g., Bernardin, 1977; Burnaska & Hollmann, 1974). However, in a study which examined the effects of rater participation in scale construction on various psychometric characteristics, with respect to two different rating-scale formats, Friedman and Cornelius (1976) concluded "participation led to more desirable psychometric operating characteristics using *either* scale format" (pp. 214-215). Nevertheless, reductions in halo error using BARS were not consistent across studies, irrespective of format (Borman, 1975; Borman & Dunnette, 1975; Bradley & Pursley, 1987).

BARS studies which attempted to limit leniency error were similarly equivocal. As was the case when examining halo error, results of leniency error investigations were inclined to compare BARS with other rating techniques. Specifically, it was claimed that summated scales and other 'profession-specific' scales (e.g., those used to assess naval officers, see Borman & Dunnette, 1975) exhibited higher levels of leniency error (Borman & Dunnette, 1975; Campbell et al., 1973). However, such results were not supported in similar studies (Bernardin, 1977; Burnaska & Hollmann, 1974). Bernardin, Alvares and Cranny (1976) concluded there was less leniency error for summated scales, provided the scale was item analysed. Despite the contradictory findings, Landy (1985) noted that by reducing the degree of ambiguity of scales (and thus improving the operational definitions of each dimension), leniency error could be overcome. Similarly, this strategy helped to reduce halo error, as the rater is less likely to give overall impressions if dimensions are defined clearly (Landy, 1985).

As a synopsis, Landy et al. (1991) concluded "that BARS have frequently demonstrated acceptable psychometric integrity" (p. 649). This is despite the integrity of BARS instruments, with respect to reliability, validity, halo error, and leniency error being ambiguous. Results are conflicting and substantive conclusions rare. The apparent discrepancies seem to emanate from the wide variety of procedures used to establish the psychometric criteria and circumstances in which the criteria were assessed. An example of this is the establishment of reliability during development, as opposed to implementation, phases of BARS instruments. Clearly, consistency in approaches to measuring these criteria needs to be established.

Although the suitability of BARS for assessment purposes is debatable, it is argued that BARS should not be so narrowly utilised, i.e., wider 'spin-off' applications of BARS warrant consideration (Blood, 1974; Jacobs et al., 1980; Schwab et al., 1975). Essentially, the spin-off applications are a result of development procedures inherent in scale construction, particularly with respect to the use of profession-related experts (Landy, 1985). Accordingly, the following discussion provides an overview of the utility of BARS, and subsequent applications of BARS data in applied settings.

Applications of BARS

Previous discussion detailed how BARS provided occupational criteria for assessing work performance. At the same time, it became apparent BARS could supply individuals and organisations with viable and useful data which fostered applications over-and-above what was ordinarily demanded of assessment criteria. Four applications stand out, and all are implicitly related to the breadth of occupational skills manifestly identified in BARS development.

The first of these is when BARS is used as a data gathering exercise for the purpose of *occupational analysis*. Underpinning this application is the assumption that most occupations are somewhat complex, and an interaction of many skills is required for effective occupational performance. This complexity was acknowledged by Smith and Kendall (1963), and they asserted that BARS identified all facets of an individual's performance through substantial analysis of successful performance. This was achieved not only through rigorous developmental procedures, but via the use of profession-related employees from various organisational levels (Schwab et al., 1975). Detailed occupational descriptions were further clarified as the "use of critical incident type of information specifies concrete behaviors that must be performed by an individual on the job" (Blood, 1974, p. 513).

The second application, a capacity to *provide detailed and accurate feedback* to employees and employers regarding job performance, was viewed in the literature as one of the prime spin-off applications of BARS (Anshel, 1995; Hom et al., 1982; Shirom, 1988). The application is grounded by the diagnostic capability of BARS data, sustained through the extensive occupational analysis procedures central to BARS methodology. Individuals evaluated their level of performance against actual occupational tasks, as examples of work were stated clearly and unambiguously. This, in turn, reflected real and experienced work situations (Shirom, 1988).

Equally, the accuracy of feedback derived from BARS data accrued several other benefits to the individual, including:

- the ability of raters to document their review with specific incidents (Zedeck & Baker, 1972), thus refining the interpretability and meaningfulness of feedback to the recipient (Hom et al., 1982);
- the high credibility of the scales, amongst workers, due to extensive job analysis (Hom et al., 1982); and,
- the establishment of specific goals for improvement by indicating explicitly which aspects of occupational performance were desirable, and those that should be avoided (Hom et al., 1982).

Organisational *diagnosis and development*, facilitated through feedback provided by BARS, presented as the third significant application (Jacobs et al., 1980). The inherent diagnostic nature of BARS data can provide the foundation for individualised, purpose-designed, training programs (Anshel, 1995; Blood, 1974; Shirom, 1988). This provides opportunities for occupational training programs to be clearly focused (Blood, 1974; Campbell et al., 1973). More generally, items collected during BARS development permits the identification of employee weaknesses 'across the board' on the same performance dimension.

The final application to note is the potential for BARS to *discriminate among various organisational levels*. Importantly, this application needed to be viewed in terms of 'performance perception' at each level of an organisation. Zedeck, Imparato, Krausz and Oleno (1974) found that, as a function of organisational level, a number of raters varied in their rating scores. This implied that various organisational levels, of any organisation, may have differing perceptions on what performance criteria are, and are not, important in an occupation. These considerations needed to be addressed when performance judgements are made by decision makers.

Blood (1974) hypothesised these four applications of BARS were previously overlooked due to its potential as a performance scaling technique. However, the four applications described above demonstrate that BARS, in combination with associated developmental procedures, need not be confined to assessment purposes only. As Blood (1974) noted, the wider applications of BARS can provide the researcher with useful information in other areas of organisational operation.

Summary

From the evidence provided thus far, it can be concluded that BARS, when used strictly as a rating instrument for assessing individual performance, is open to question. It has not been unequivocally established that BARS provides reliable and objective measures of an individual's work performance. Despite these concerns, positive views on BARS methodology predominate (Stoskopf et al., 1992). This is primarily a consequence of BARS

procedures almost guaranteeing that resulting scales are constructed carefully (Landy, 1985). Hence, it has proved to be a reliable methodology for the determination of relevant and meaningful occupational criteria. Accordingly, it is argued that the methodology should be viewed from a broader perspective than assessment purposes *per se* (Blood, 1974; Jacobs et al., 1980).

This latter perspective of BARS methodology provides a framework for discussion in the following section. Specifically, this section traces the development of 'hybrid' BARS methodology, which has recently been employed successfully to identify occupational performance criteria for sports officials and other occupations.

THE EVOLUTION OF HYBRID BARS

Aside from the psychometric concerns identified about BARS in the previous section, the application of traditional BARS methodology raised three other pertinent issues. Firstly, the tight focus on behavioural aspects of performance was claimed to be impractical and unwieldy (Masters & McCurry, 1990). Instances have been cited where occupational task analysis became too detailed to be practical, e.g., lists of 4300 discrete behaviours to drive a car, and 3001 functions performed by a teacher (Masters & McCurry, 1990). This approach tended to 'atomise' the profession and ignore other important performance attributes. As concluded by Carmichael (1993), "narrowly focused and rigidly demarked skills ... are of a past that is receding" (p. 17).

Moreover, Preston and Walker (1993) maintained that traditional competency approaches did not give significant recognition to performance attributes such as knowledge, abilities, and attitudes (Gonczi, Hager, & Oliver, 1990; Hager & Beckett, 1995). It was claimed the relationship and of these attributes to performance were too often overlooked despite their essential contribution for the "richness of work to be captured" (Hager & Beckett, 1995, p. 3).

Secondly, the nature of BARS procedures meant the development of performance dimensions and behavioural examples were often time consuming, tedious, and expensive (Borman, 1975; Jacobs et al., 1980; Landy, 1985; Landy & Farr, 1980; Leap & Crino, 1993; Shirom, 1988). Consequently, practitioners were cautious in embracing BARS for logistical reasons alone.

Lastly, extensive research during the 1970s into BARS, as an assessment tool, made further research superfluous (Jacobs et al., 1980). It is maintained that to continue research along the assessment theme is to uncover nothing new, and would result in a disservice to the methodology (Jacobs et al., 1980).

Although these three issues are not linked through any substantial theoretical underpinning, nonetheless, they indicate that BARS, in its traditional form, warrants reappraisal. As Jacobs et al. (1980 p. 635) acknowledged:

What is desperately needed is a more far reaching approach to performance ... one which includes scale construction and utilization as a starting point, incorporates performance feedback, and culminates in organizational efforts towards training ...

Subsequently, a hybrid version of BARS methodology evolved during the 1980s (for example, see Anshel et al., 1987), with development continuing during the 1990s (Anshel, 1995; Anshel & Webb, 1991; Jessup, 1994; Moore & Webb, 1995; Webb, Jessup, Moore, & Landy, 1994). Although relatively new, it does show potential in overcoming concerns associated with traditional BARS. Moreover, the hybrid version – while still maintaining a behavioural predilection – incorporates aspects of the ‘integrated’ or ‘holistic’ view of competence described previously, where attributes of the worker are considered also in determining performance requirements (Hager & Beckett, 1995; Preston & Kennedy, 1995; Preston & Walker, 1993).

Specifically, the BARS variant incorporates a number of operational routines characteristic of traditional BARS methodology, with scale development embracing goal analysis procedures developed by Mager (cited Cyrs, Dobbert, & Grussing, 1976). These routines are outlined in the following discussion. Particular examination focuses on major deviations from traditional BARS, and the related validity and reliability of hybrid BARS instruments.

Development of Hybrid BARS

Similar to traditional BARS methodology, the hybrid BARS procedure utilised a predetermined framework. This too has undergone minor alterations since its inception. However, as described by a number of sources (Anshel, 1995; Anshel et al., 1987; Anshel & Webb, 1991; Jessup, 1994; Moore, Webb, & Dickson, 1997; Webb et al., 1994), the hybrid methodology involved essentially five discrete and sequential steps. Like traditional BARS, these were designed to ensure rigorous identification of skills and performance dimensions. The steps are outlined below, and represent a synthesised overview (as described by the above sources) of the methodology.

Step 1. The selection and meeting of profession-related experts. Profession-related experts are responsible for the generation of occupational performance criteria relevant to the role under investigation. They are selected according to a similar criteria prescribed in traditional BARS literature, i.e., familiarity, knowledge and/or expertise in the profession or role under investigation. Experts are brought together for the expressed purpose of identifying profession-specific performance criteria.

Step 2. *Meeting to generate profession-specific performance domains and competencies.*

The profession experts meet for a one-day period. They progress through a number of clearly defined procedures, the net result being a list of performance domains and competencies essential to effective occupational performance. However, before entering into meeting procedures, the cohort of experts are divided into two (and sometimes three) identical sub-groups or 'panels.' This operation ensures expertise is reflected across panels. Additionally, each panel is required to operate independently of other panels in working through the procedures. Specific meeting procedures include:

- a. each group member *identifying descriptive phrases or words* which are representative of the occupation or role under investigation (Anshel, 1995; Cyrs et al., 1976). Panel members come together to combine and edit the list by eliminating duplicate or non-essential entries. It is not uncommon for new items to be added, modified or rewritten (Anshel & Webb, 1991; Cyrs et al., 1976). This procedural phase is completed when the panel reaches consensus on a composite list of descriptive words and phrases.
- b. panel members *classifying descriptive words and phrases into related or homogenous categories* (Jessup, 1994). This task is undertaken by panel members working independently. The categories developed by each expert are analogous with the performance domains described previously in traditional BARS methodology. Subsequently, panel members again come together to reach a consensus about the relevancy of the performance dimensions. Only those dimensions that reached 100% agreement are retained for further development (Cyrs, 1979).
- c. the *development of specific competencies* within each performance dimension. This is achieved by expanding the descriptive words and phrases developed previously (i.e., point 'a', and grouped subsequently in point 'b') and re-writing them to describe an action, behaviour, or trait that is observable or measurable.

The collective noun 'competencies', given to these actions, behaviours, and traits, is not predicated on BARS literature. This is primarily because the literature does not present consistent terminology. For example, Anshel (1995) called each performance dimension a 'competency', with specific skills referred to as descriptors (i.e., skills describe a competency). Alternatively, Moore et al. (1997) used 'competency' in the same manner as described in the present study. This latter approach was grounded in contemporary literature on occupational competencies where *specific* descriptions of occupational

performance were designated the collective noun 'competency' (Murphy & Cleveland, 1995; Preston & Kennedy, 1995; Rutherford, 1995)

Step 3. *Amalgamation of lists.* Panel listings of performance dimensions and competencies are amalgamated to form a single structure. This process is not usually conducted by panel members, but by nominated 'panel leaders' and/or 'process observers' (Anshel, 1995; Anshel & Webb, 1991). Similar dimensions and competencies developed by each panel are identified and combined. The procedure is viewed by Anshel (1995) as analogous to the *retranslation* process described by Dickinson and Trice (1973; 1977) in traditional BARS development.

Step 4. *Verification of combined list.* A draft list of the combined performance dimensions and competencies is distributed to all panel members for comment. Panel members are required to check the list for content and terminology. This phase is designed to overcome any amalgamation problems that may have occurred in Step 3 (Cyrs et al., 1976). Only items that reached 90% agreement are retained (Anshel, 1995).

It is at this point that performance dimensions and competencies are fully developed. Furthermore, this point in the process represents the conclusion of input by members of the expert panels.

Step 5. *The external validation of performance dimensions.* In an additional step not undertaken normally in traditional BARS methodology, a five-point scale is placed next to the performance dimension (Anshel, 1995; Anshel & Webb, 1991). The list is submitted for "external validation" (Anshel, 1995, p. 16) by sending the instrument to a wider representative sample of practitioners. This group is asked to rate the importance of each dimension on a 1-to-5 Likert scale. Dimensions which meet with 90% agreement about its importance are retained (Anshel, 1995; Anshel & Webb, 1991; Moore & Webb, 1995). Implicit in this step is the establishment of face validity (Anshel, 1995), an important pre-requisite if dimensions are to be used as a basis for assessment and training procedures.

A feature of the five steps outlined above is that they overcome a number of the logistical problems associated with traditional BARS methodology. In particular, there are savings in time and money. This is evidenced through two specific examples. Firstly, the meeting of experts which generates performance dimensions and competencies is completed within one full day (Step 2), and consequently, costs are minimised. Secondly, Phases 4 and 5 can be undertaken via posting, and can be completed in a matter of weeks. Overall, the five-step procedures provide a more succinct process for identifying essential performance criteria. Moreover, the tedious procedures inherent in traditional BARS are avoided. Importantly,

these advances are possible while still maintaining diversity of opinion and rigorous item selection.

However, of concern is the relative infancy of the hybrid process. This is evident through the volume, and pedigree, of supporting research (which is not commensurate to traditional BARS). Particularly, issues surrounding procedural differences between BARS methods require closer scrutiny.

Issues Surrounding Hybrid BARS Methodology

The hybrid methodology incorporates many of the fundamental procedures and assumptions underlying traditional BARS methodology. However, this technique made a number of alterations, the most obvious being: the manner of selecting performance dimensions and competencies (specifically, their retention and character), and the use of Likert scales to anchor dimensions (as opposed to the use of a vertical continuum of descriptors). Interestingly, these changes were recognised in hybrid studies (see for example, Anshel, 1995), however, no stated or latent reason for change was provided by the literature. This latter point notwithstanding, each issue is considered separately below.

The first difference is the *retention* of all performance dimensions and competencies in the hybrid model (see Step 3). Hybrid BARS ensures no performance criteria are eliminated unless duplication occurs across panels. A possible benefit of this adaptation is that it may help overcome the inter-rater reliability problem manifest in traditional BARS methodology (by way of revision, traditional BARS eliminated items via the *retranslation* and *scaling* phases, thus leading to assessors observing worker-behaviours that were not listed in the scales).

This benefit notwithstanding, the retention of all performance criteria raises a fundamental question, namely, do expert panels identify comprehensively the essential and important performance criteria of an occupational role? This question is answered, in part, through two related operational steps central to hybrid BARS. Firstly, the retention of all performance dimensions and competencies ensures the entire performance domain of the role under investigation is covered thoroughly. All descriptions of work are included, thus minimising the possibility of essential criteria being eliminated due to arbitrary criteria. Secondly, the *external validation* process (Step 5) requires a wide sample of respondents to assess the importance of each performance dimension generated by the expert panels. Dimensions are discarded only if obvious disagreement existed within the larger sample.

With respect to this latter endeavour, hybrid BARS has been relatively successful. Anshel et al. (1987) reported that only two of 22 dimensions developed by expert panels for high-school football coaches were seen as 'not essential' for successful coaching. Subsequent

studies reported similar success with this procedure. Impressively, *all* performance dimensions developed for respective studies on Touch referees (Anshel & Webb, 1991), classroom management (Jessup, 1994), and basketball referees (Anshel, 1995) were accepted as important and/or essential by wider populations of practitioners.

The main conclusion to be drawn from these studies (Anshel et al., 1987; Anshel & Webb, 1991; Jessup, 1994) was that hybrid BARS methodology did produce an extensive and valid range of occupational performance dimensions. However, it should be noted that these studies had the performance dimensions 'externally validated' only. Success in generating accurate and meaningful competencies was not assessed. Consequently, questions remained concerning the capacity of hybrid BARS to identify comprehensively essential competencies which were not only reflective of performance dimensions, but also indicative of effective performance in their own right.

The second significant difference between hybrid and traditional BARS related to the character of competencies attached to each performance dimension. In traditional BARS instruments, performance dimensions are described by items which represent effective, average and poor occupational performance. However, hybrid BARS dimensions contain competencies which reflected desirable performance criteria only, with all descriptors being of equal value (Anshel, 1995).

Reasons for this change are unclear. However, it was indicated that traditional BARS formats did not lend themselves to scale development in the sporting context (Anshel, M. 1998, pers. comm., 20 January). Anshel argued that it was very difficult for panel members to agree on an hierarchical rating of competencies, i.e., the issue of which competency is superior to another was so contentious that it was impossible to reach a consensus (Anshel, M. 1998, pers. comm., 20 January). This experience is similarly reflective of problems encountered in studies by Landy and Guion (1970), and Harri and Zedeck (1973) using traditional BARS methodology. They too found the determination of behaviours that reflected poor-to-mid-range performance difficult.

A second reason why the traditional hierarchical structure of BARS may have been altered (although no such reason is given in the literature) was early hybrid BARS was not used for rating and assessment *per se* (e.g., Anshel et al., 1987; Anshel & Webb, 1991). In the main, BARS was used in these studies to identify the important and essential performance criteria of respective roles and professions (this is indicative of 'spin-off' applications suggested for traditional BARS procedures, see Blood, 1974; Borman & Dunnette, 1975). Consequently, it can be hypothesised that the need for a continuum of skills and behaviours which reflect effective-to-poor performance was superfluous.

The final alteration under review centres on the adoption of Likert scales to anchor performance dimensions. Interestingly, foundation BARS literature maintained that BARS was developed to overcome some of the psychometric problems inherent in summated and graphic rating scales (Smith & Kendall, 1963). However, there was subsequent evidence to suggest that the incorporation of Likert scales was acceptable. Borman (1979) argued that Likert scales, placed next to behavioural indicators, may actually improve the reliability of the instrument due to several scores per dimension (rather than one score) being used to assess performance. Moreover, other research demonstrated that the measurement properties of traditional BARS instruments were no better than well constructed scales using graphic rating scales or summated scales (Bernardin, 1977; Bernardin et al., 1976; Borman, 1979; Burnaska & Hollmann, 1974; Landy, 1985).

This brings into question the validity and reliability of hybrid BARS data, two aspects of the newer technique which have not been investigated extensively. However, one study did take steps to address these two measurement criteria. In a study of basketball officials, Anshel (1995) investigated issues surrounding the validity and reliability of a hybrid BARS instrument, the results of which form the basis of the following discussion.

The validity of the Anshel's instrument was examined through content, construct, and predictive validity. Anshel (1995) asserted that content validity was established through the use of occupational experts in the developmental phase of BARS. While this premise is sustainable, closer scrutiny of the acceptance of competencies, as opposed to just performance dimensions, would add credence to this form of validity. As noted previously, the external validation process required respondents to assess the importance of dimensions only. The success of BARS in developing accurate competencies is yet to be assessed, thus claims of acceptable content validity are premature.

Aside from this concern, Anshel (1995) used an 'external validation' process as a form of 'de facto' factor analysis to assess the placement of competencies into dimensions. In his study, respondents indicated their agreement or disagreement that each of the 56 descriptors belonged to its designated performance dimension. As a result, all descriptors were accepted as representative of respective competencies. Although impressive at face value, this form of categorisation is questionable. It is arguable that respondents took the time, or possessed the expertise, to examine objectively all descriptors independently of the dimension structure. Further empirical investigation is required to establish the validity of performance dimension structures developed by hybrid BARS procedures.

In a related point, redefinition of the *retranslation* process by Anshel (1995) raised an additional concern. As discussed previously, Anshel described this process as the amalgamation of performance criteria, developed by independent expert groups, at the conclusion of scale construction. Moreover, Anshel claimed this process was the same as

the retranslation process followed by Dickinson and Trice (1973). Perusal of the Dickinson and Trice study does not reflect Anshel's claim. In fact, the study of Dickinson and Trice followed essentially the same retranslation process as described by Smith and Kendall (1963). Reasons for this contradiction were not evident, so questions remain over the capacity of hybrid BARS techniques to classify accurately competencies into appropriate performance dimensions.

The second form of validity examined by Anshel (1995), i.e., construct validity, was assessed by testing the hypothesis that the instrument could differentiate between advanced and novice officials (the construct for the study was referee performance effectiveness). With skill level as the independent variable and performance dimensions as the dependent variables, a one-way MANOVA computed an alpha value of $p < .001$. Anshel maintained that this significant result demonstrated clearly the instrument's ability to discriminate between novice and highly skilled officials, thus construct validity was affirmed.

However, the establishment of construct validity in this manner is tenuous. Construct validity is more abstract than other types of validity, and, therefore, more difficult to validate. Nevertheless, it is accepted generally that construct validity measures hypothetical traits and concepts such as intelligence, motivation, and authoritarianism (Borg & Gall, 1989; Crowl, 1993). In the study by Anshel (1995), the selection of 'referee performance' as a construct is questionable at best. Analysis of underlying traits, manifest within the behavioural indicators, as implemented by Stoskopf et al. (1992) may have improved the rigour of Anshel's findings.

Lastly, stepwise discriminant function analysis was employed to assess predictive validity. This analysis indicated that 11 of the 13 performance dimensions significantly predicted group membership ($p < .05$ to $p < .001$). Anshel (1995) claimed that these validity findings confirmed the hypothesis that the BARS instrument discriminated adequately between novice and experienced officials. Such a claim is sustainable in this context.

Aside from validity findings, Anshel (1995) also reported the reliability of scales. This was determined via two methods. Firstly, the degree of consistency for descriptors within each performance dimension was computed using Cronbach's alpha. Co-efficients ranging from 0.79 to 0.93 were reported, thus showing that the examples which described each performance dimension were consistently related to each other (Anshel, 1995). Secondly, Pearson's product-moment correlations were used to assess inter-rater reliability for all observations. A co-efficient of 0.88 was recorded, showing high consistency between the ratings provided for each game (Anshel, 1995).

In overview, it is obvious that the overall body of evidence relating to the validity and reliability of hybrid BARS is still sparse. With only one study examining these issues with

any rigour, definitive conclusions are tenuous, particularly with respect to the accurate selection of performance descriptors and construct validity. However, as a whole, the published results are promising, and provide encouragement for the future use of the modified BARS technique.

CONCLUSION

BARS methodology had its genesis in the need to assess accurately occupational performance (Smith & Kendall, 1963). In doing so, it was seen to provide a solution to many of the problems inherent in previous assessment methods, particularly those relating to the validity and reliability of assessment instruments. However, in this capacity, the traditional promises and expectations of BARS were not emphatically realised. Notwithstanding unfulfilled potential, there were some obvious and useful 'spin-off' advantages from the methodology (Blood, 1974; Jacobs et al., 1980). These included the ability of BARS methodology to guarantee meaningful occupational analysis, and to provide accurate descriptions of effective occupational practices.

Following encouragement to review the focus of BARS research (Jacobs et al., 1980), a hybrid BARS methodology emerged. In its earlier applications, the methodology focused predominantly on using BARS for occupational performance analysis, although one later study applied hybrid BARS data for assessment purposes as well (see Anshel, 1995). The new methodology was seen to overcome many of the logistical problems associated with traditional BARS methodology, while concurrently presenting researchers with a procedure that has many of the benefits of traditional BARS methodology, e.g., the ability to produce performance criteria that were viewed consistently as essential and important for effective occupational performance. In doing so, the hybrid version has also the capacity to incorporate employee attributes into performance models.

Moreover, the hybrid methodology is gradually establishing itself as an effective method to develop performance rating instruments, although more research is needed concerning the development of performance domains and descriptors before definitive conclusions can be made. Two issues, in particular, require substantiation. Firstly, the presumption that panel experts identify accurately specific occupational competencies is untested. Secondly, the classification of competencies, into performance dimensions, is subjective and intuitive. For hybrid BARS to be seen as a credible procedure for developing accurate descriptions of occupational performance, these two issues require empirical investigation and assessment.

RESEARCH QUESTIONS, HYPOTHESES AND THEMES

From the issues raised in the review of literature in Chapter 1, and the unresolved issues surrounding hybrid BARS procedures described in this chapter, a series of themes, questions, and hypotheses arose which seemed worthy of further empirical investigation.

In their most general form, these themes are:

- Theme 1. What occupational performance criteria are essential for elite soccer referees?
- Theme 2. How suitable is the hybrid BARS technique for identifying performance criteria of elite soccer referees?
- Theme 3. How are refereeing performance criteria perceived by various soccer stakeholders such as referees, assistant referees, referee inspectors, players, and coaches?
- Theme 4. Do soccer stakeholders (as identified in Theme 3) hold different perspectives about the role of referees?
- Theme 5. When given the opportunity to respond openly about the nature of elite soccer refereeing in Australia, what information is provided by soccer stakeholders?

In order to focus more clearly on each theme, a set of specific research questions and, where appropriate, hypotheses were developed. These are provided below.

Theme 1

What occupational performance criteria are essential for elite soccer referees? Implicit in this theme is the identification of a variety of refereeing competencies, and their classification into performance dimensions.

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

Theme 2

How suitable is the hybrid BARS technique for identifying performance criteria of elite soccer referees? For hybrid BARS to be embraced as a valid research methodology, the integrity of performance domains and competency development needs to be confirmed, particularly if items are to form the basis of research instruments.

- Question 2.1 Do major soccer stakeholders support the importance of refereeing competencies identified through the hybrid BARS process?
- Question 2.2 Are performance dimensions, identified by occupational experts, supported using statistical analyses?
- Question 2.3 Do hybrid BARS instruments demonstrate acceptable reliability characteristics?

Theme 3

How are refereeing performance criteria perceived by various soccer stakeholders such as referees, assistant referees, referee inspectors, players, and coaches? Notions embedded in this theme include the relationship between criteria seen to be important, and the quality of their execution. The theme is addressed through three research questions and one related hypothesis.

- Question 3.1 How are essential refereeing performance criteria perceived by soccer stakeholders in terms of relative importance?
- Question 3.2 How are essential refereeing performance criteria perceived by soccer stakeholders in terms of relative preparedness?
- Question 3.3 Is it perceived that referees execute the most important criteria with high levels of preparedness? This research question can be re-expressed in the form of the following hypothesis, thus open to empirical testing:
- Hypothesis 3 There will be no significant difference between the relative importance of elite soccer refereeing criteria and the preparedness of referees for undertaking these criteria.

Theme 4

Do soccer stakeholders (as identified in Theme 3) hold different perspectives about the role of referees? While it is possible to view players, coaches, and officials of soccer as an homogenous group, they do present different approaches and outlooks on the game. Hence, possible differences and similarities in their views may offer insight into the performance requirements and standards of refereeing work.

Specifically, do players, coaches, referees, assistant referees, and referee inspectors perceive the same relative *importance* of refereeing performance criteria? Also, do players, coaches, referees, assistant referees, and referee inspectors view the *performance* of these

criteria similarly? These two questions can be re-expressed in the form of the following hypotheses:

- H 4.1 There will be no significant difference of opinion about the relative *importance* of elite refereeing performance criteria among referees assistant referees, referee inspectors, players, and coaches.
- H 4.2 There will be no significant difference of opinion about the relative *preparation* of elite refereeing performance criteria among referees, assistant referees, referee inspectors, players, and coaches.

Theme 5

When given the opportunity to respond openly about the nature of elite soccer refereeing in Australia, what information is provided by soccer stakeholders? As BARS is a closed system, there is a limit to the amount of information that can be provided by respondents. Specifically, there is no provision to comment on any individual, or collective, aspect of elite refereeing that respondents may feel strongly about. The purpose of this theme is to address such limitations through the analysis of qualitative comments related to the following questions.

- Question 5.1 Within the context of a free response, how is the current standard of refereeing perceived?
- Question 5.2 Which aspects of refereeing performance were the focus of most comment?
- Question 5.3 How do qualitative comments substantiate, and elaborate upon, the quantitative findings?

In summary, the research themes, questions and hypotheses are concerned with 'illuminating' an under-researched area. Each theme is of intrinsic value and interest, yet also has the potential to clarify further the nature of the referees' role and the diversity of skills required of referees. However, in addressing these themes, a more general form of inquiry arises, i.e., what are the implications for the training and development of elite soccer referees. Although these implications are not expressed as specific themes or questions, they are nonetheless worthy of discussion, and as such are addressed within the context of this investigation.