Journal of Sport Management, 1996, **10**, 446-462

Factors Considered Important for Evaluating Canadian University Athletic Coaches

Joanne MacLean
University of Windsor

Dorothy Zakrajsek
The Ohio State University

Evaluating coaching performance based on the use of job-specific assessment criteria has been particularly problematic for college athletic departments. The purpose of this study was to assess the importance attached to six dimensions of criteria rated by administrators ($n = 87$) and coaches ($n = 532$) in the Canadian Interuniversity Athletic Union. The six dimensions were team products, personal products, direct task behaviors, indirect task behaviors, administrative maintenance behaviors, and public relation behaviors. The results of Multivariate Analysis of Variance (MANOVA) and repeated measures ANOVA showed that, in general, administrators and coaches held similar beliefs about the criteria important for coaching evaluation, but they had some differences in the order of importance of the dimensions. Both groups rated direct task behaviors—that is, the specific abilities or skills used directly in the day-to-day practice of coaching—as the most important dimension of grouped evaluation criteria.

Employee performance appraisal is central to the field of organizational behavior (Vecchio, 1988). As such, it is a necessary process for all jobs, regardless of level or complexity. The appraisal of work performance assesses the human factor in organizational effectiveness and has been defined as ". . . a process of formally evaluating performance and providing feedback on which performance adjustments can be made" (Schermerhom, Hunt, & Osbom, 1985, p. c-2).

Evaluation research has historically been dominated by a global organizational focus that investigates products, processes, and entire organizations (Murphy & Cleveland, 1991), although more recently, greater attention is given to evaluating personnel because of the direct relationship it has to successful organizations. Accordingly, the ultimate success of an organization is predicated upon the quality and performance of personnel (Cummings & Schwab, 1973; Darling-Hammond, Wise, & Pease, 1983; Murphy & Cleveland, 1991; Wexley & Latham; 1981).

Research efforts by scholars in fields such as education, business, and sociology provide much about the diverse nature of performance appraisal (Beer, 1987;

Joanne MacLean is with the faculty of Human Kinetics, University of Windsor, Windsor, ON, Canada N9B 3P4. Dorothy Zakrajsek is with the School of HPER, the Ohio State University, Columbus, OH 43210-1284.

Darling-Hammond, 1986; Lane, 1990; Lawler, Mohrman, & Resnick, 1984; Peterson & Comeaux, 1990; Rieder, 1973). Some of their investigations examined evaluation methods and purpose and sought to explain the power and politics therein. Some promoted fixed standards for judging performance, some considered environmental and organizational influence, and others concentrated on measuring rating accuracy and the motivation of the rater. In summary, generic models sought answers to the questions: "What should be evaluated?" and "How should the evaluation be conducted?"

Although it is very important to articulate the purpose, timing, method, standards, and level of analysis of the evaluation, it is crucial to first define performance by giving substance to the specific criteria needed in the assessment process (Kurtz, Mueller, Gibbons, & Dicataldo, 1988). There is little utility in accurately measuring the job performance of an individual if the criteria measured are unrelated or inconsequential to the job. Thus, the foundation criteria must be represented in a formalized procedure set in advance.

Jobs are created to fulfill specific purposes or certain job functions within the framework of organizational objectives. These responsibilities define the job description which, in turn, becomes the focus of performance appraisal. Buford, Burkhalter, and Jacobs (1988) and many other authors (Deets & Tyler, 1986; Ilgen & Bames-Farrell, 1984; Murphy & Cleveland, 1991; Torrington, Weightman, & Johns, 1989) have argued the need to link job descriptions to performance appraisals. Historically, the practice of aligning evaluation with actual job requirements has been problematic.

Generally, job-specific performance appraisal is broadly recognized as an essential component for valid evaluation. Murphy & Cleveland (1991) argue for more than job descriptions by making, the case that contextual factors of the environment would also be considered:

> Performance appraisal cannot be adequately understood outside of its organizational context—the same appraisal system, the same criteria for evaluating ratings, the same rater-training programs, and so on are not the same if they exist in different contexts. (p. 25)

The authors defined contextual factors as a heterogeneous mix of factors that range from the social and legal system to the climate and culture within the organization. Therefore, the argument is made that performance appraisal should be based on the job description, with consideration given to contextual factors derived from the environment.

Ideally, job descriptions are specific enough to help define job performance, but this is not always the case. Job descriptions are often too vague to be of much use in evaluating performance (Buford et al., 1988). In an effort to add specificity, job performance has been defined in terms of (a) behaviors indicative of performing a job (e.g., the assembly line worker's accuracy in assemblage, dependability on the job, or ability to cooperate with co-workers) or (b) in terms of the results of those behaviors (e.g., the number of toys packaged per hour) (James, 1973; Smith, 1976). Researchers have focused on assessing the utility of both categories of appraisal criteria and concluded that basing performance appraisal on the results of behaviors is too short term. Landy and Farr (1983) believe that relying exclusively on results of behaviors can lead to behaviors that are dysfunctional for the organization. Murphy and Cleveland (1991) suggest that too much emphasis on results

makes it difficult to determine what is being evaluated, the person or the situation:

> Results-oriented criteria can lead supervisors and subordinates to ignore a wide range of behaviors (e.g., maintaining good interpersonal relations) that are critical to the survival and effectiveness of the organization but are not uniquely tied to any given product or result. (p. 92)

Therefore, job performance is more appropriately defined in terms of kinds of behaviors in addition to behavioral outcomes. These different types of behaviors involve relevant task goals of the job as well as nontask goals, which help describe the domain of performance (Astin, 1964; Carroll & Schneier, 1982; Fleishman & Quantance, 1984; Salvendy & Seymour, 1973). Relevant task goals are those behaviors central to the job or to the quality of task-oriented behavior (Carroll & Schneier, 1982). Nontask goals are those behaviors included within the domain of the job but that are peripheral to the specific task (Astin, 1964). These job-related behaviors and results of behaviors set the limits for defining appraisal criteria. Murphy & Cleveland (1991) suggested that performance appraisal has traditionally been viewed as a measurement problem and that much of the research to date has focused on better, more accurate, or more cost-effective measurement techniques. However, effective measurement techniques are conditional to job-specific performance criteria, and the development of such criteria presents a major challenge to researchers and practitioners alike.

In an effort to provide direction for the development of job-specific performance criteria, MacLean (1993) generated a procedural model for the purpose of responding to the question, "What criteria should be utilized for the appraisal of an individual in a specific job?". The literature-based theoretical model (Figure 1) shows a process for developing independent evaluation criteria for use in specific performance evaluation procedures (MacLean, 1993, 1994). The process involves a three-step procedure: job assessment, creating the job description, and defining the domain of performance. The domain of performance is comprised of behavioral *product* factors and behavioral *process* factors; the behavioral process factors are further categorized as either task related or maintenance related. The impact of environmental factors, both internal or external to the organization, are considered. Subsequently, the behavioral product factors, the task-related behavioral process factors, and the maintenance-related behavioral process factors were each divided into two subcategories to yield a two-dimensional model of the dimensions of coaching performance. This is schematically shown in Figure 1 and operationally defined in Figure 2.

Assessing employee performance is both an expectation and a duty for employing organizations. In higher education, institutional efforts to develop fair and representative appraisal systems for faculty and staff is an ongoing process (Elmore & LaPointe, 1975; Kurtz, Mueller, Gibbons, & Dicataldo, 1988; Ynapp, 1982). However, in our opinion, athletic departments have not kept pace with the larger institution. While evaluation in such departments may be well-intentioned, the process for assessing coaching performance seems weak. According to Leland (1988) coaches often receive little evaluative feedback other than game day results. Bennice (1990) believes that coaching evaluation has remained an enigma due to difficulty with collecting accurate data and implementing the evaluative process. A recent study of Canadian universities showed that coaching evaluation was often performed through informal procedures and undocumented criteria (MacLean

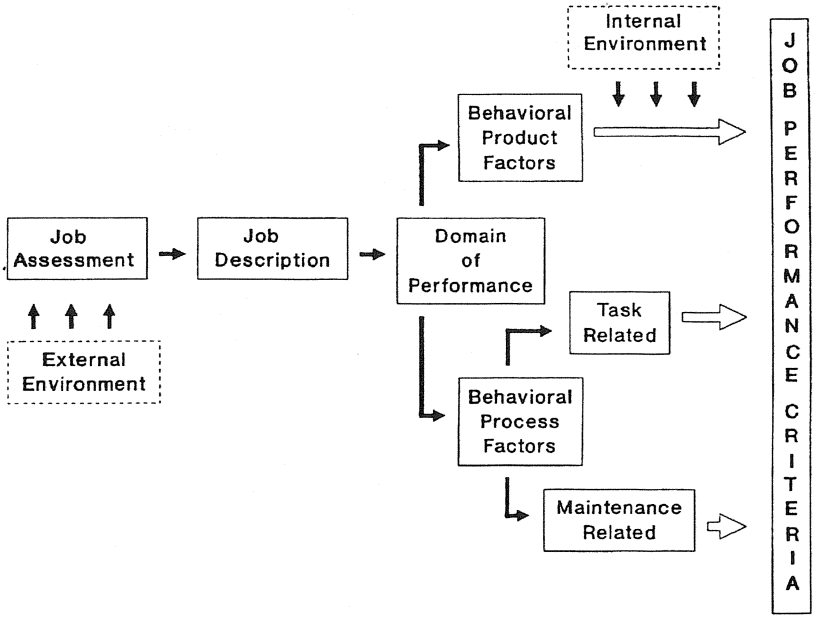

Figure 1 — Comprehensive model for the development of performance appraisal criteria.

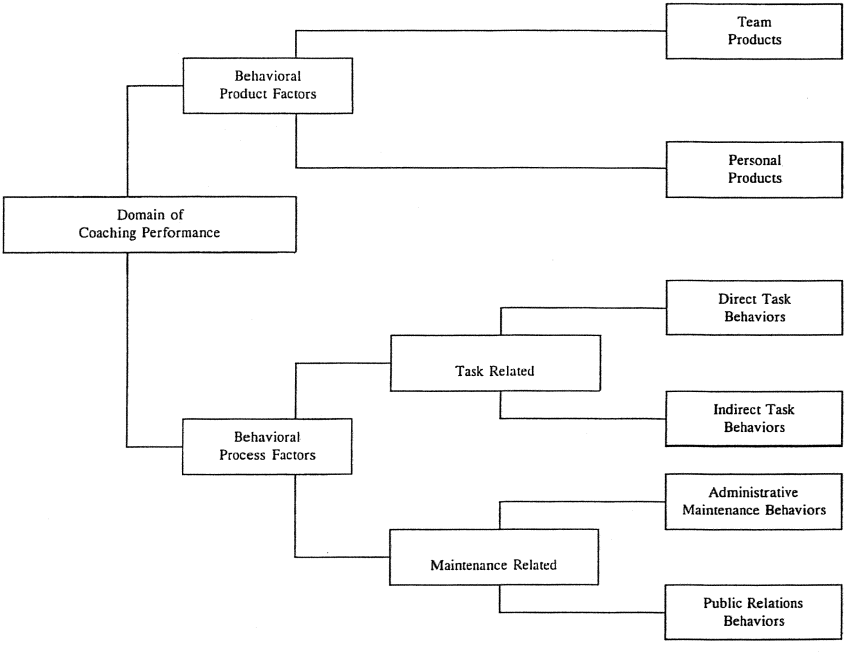

Figure 2 — Theoretical model of dimensions of coaching performance.

& Zakrajsek, 1994). Other issues related to purpose, procedures, criteria, and type of coaching contract have contributed to inadequate evaluations of college coaches (Anderson, 1985; Barber & Skoglund, 1981; Bennice, 1990; Leland, 1988; MacLean & Zakrajsek, 1994).

Part of the neglect stems from difficulty in identifying appraisal criteria. In addition to game and practice, coaching consists of a myriad of responsibilities, including recruiting, program administration, scouting opponents, public relations, and so forth (e.g., Norcross, 1986; Sabock, 1985). Research on the traits of successful coaches uncovered factors such as commitment, cooperation, loyalty, and enthusiasm (Knorr, 1989). Attempts to operationalize a set of criteria have often resulted in fragmented lists. Margolis (1979) suggested that more objective criteria are needed to evaluate overall coaching performance. Leland (1988) proposed that evaluation criteria should emanate from the coach's written job description, whereas others hold that coaches are teachers, making a case to use for coaches similar evaluation processes and criteria as those used for instructors (Martin, Arena, Rosencrans, Hunter, & Holly, 1986).

The foregoing literature strongly suggests that identifying job-specific evaluation criteria is crucial to successful performance appraisal and that evaluating coaches is a process fraught with difficulty. The purposes of this study were to assess the level of importance given to identified dimensions of performance evaluation criteria for collegiate coaches through ratings by Canadian intercollegiate athletic administrators and coaches, and to test the degree to which the perceived ratings were congruent with the dimension of evaluation criteria proposed in the theoretical model. The data for coaches and administrators was also grouped by independent variables to determine if the ratings differed between gender and experience levels; for coaches, the data was also grouped by employment status and gender of team coached.

Method

Subjects

This study was designed as a full population survey of all athletic administrators (athletic directors, coordinators, and their assistants, where applicable; (n = 87) and coaches (n = 532) employed by Canadian Interuniversity Athletic Union (CIAU) institutions (n = 45). Athletic directors and coaches were identified by the current CIAU directory. Individuals who held dual positions as administrator and coach (n = 4) were designated as administrators and deleted from the coach frame. After two follow-up reminders, responses were received from 77 administrators (89%) and 363 coaches (68%). Subject data demographics are presented in Tables 1 and 2.

Instrument

Data were compiled using a questionnaire consisting of two sections—The Scale of Coaching Performance (SCP) and demographics. The SCP was based on the previously defined model for developing job-specific performance criteria and evolved over three construction stages using standard validity and reliability measures at each stage.

The development of the SCP and its psychometric properties are reported in a companion paper (MacLean & Chelladural, 1995). Briefly, the SCP (see Figure 3) consists of 35 items to measure (a) team products (n = 4), (b) personal products

Table 1 Coach Data Demographics (n = 357)

CIAU Sports		
Basketball	55/74	74%
Cross-country	11/25	44%
Field hockey	15/18	83%
Football	23/23	100%
Gymnastics	11/16	69%
Ice hockey	30/37	81%
Soccer	37/60	62%
Swimming	20/34	59%
Track & field	12/15	80%
Volleyball	47/63	74%
Wrestling	10/15	67%
Total	270/364	74%
Non-CIAU Sports		
Badmintton	8/13	62%
Curling	7/15	47%
Fencing	7/12	58%
Figure skating	5/7	71%
Golf	6/10	60%
Rowing	8/14	57%
Rugby	11/21	52%
Skiing	7/15	47%
Squash	7/16	44%
Synchro	3/6	50%
Tennis	7/13	54%
Waterpolo	11/13	85%
Total	87/142	61%
Sex		
Female	88/118	75%
Male	269/414	65%
Employment		
Full-time	202	
Part-time	155	
Years of univ. coaching exper.		
1–4	154	
5–9	94	
10 yrs & over	109	
Team coached		
Women's	125/181	69%
Men's	154/245	63%
Combined	78/115	68%

Table 2 Administrator Data Demographics (n = 77)

Position		
Athletic Director/Department Head	39/45	87%
Athletic Coordinator/Assistant	38/42	91%
Total	77/87	89%
Sex		
Female	20/25	80%
Male	57/62	92%
Univ. Athletic Admin. Exper.		
1–4 yrs	20	
5–9 yrs	16	
10 yrs & over	41	

(n = 4), (c) direct task behaviors (n = 10), (d) indirect task behaviors (n = 5), (e) administrative maintenance behaviors (n = 8), and (f) public relations behaviors (n = 4). The internal consistency estimates ranged from .67 to .87 for a mean of .78 in the data on athletic administrators, and from .65 to .87 for a mean of .76 in the data on coaches.

Scale of Coaching Performance

Dimension	Operational Definition
D1. Team Products	Outcomes of coaching that accrue only to the team or individual athletes comprising it.
D2. Personal Products	Outcomes of coaching that accrue only to the coach.
D3. Direct Task Behaviors	Application of interpersonal skills and appropriate strategies and tactics in enhancing the performance of individual athletes and the team as a whole.
D4. Indirect Task Behaviors	Activities such as recruiting, scouting, application of statistics that contribute indirectly to the success of the program.
D5. Administrative Maintenance Behaviors	Adherence to policies, procedures, and budget guidelines, and interpersonal relations with superiors and peers that strengthen the administration of the whole enterprise.
D6. Public Relations Behaviors	Liaison activities between one's program and relevant community and peer groups.

Figure 3 — Operational definitions of the scale of coaching performance.

The 35 items were preceded by the question: How important is each of the following criteria in evaluating the job performance of a university coach? The order of the criteria was chosen at random (five different versions were used) and appeared opposite a seven-point Likert-type scale ranging from Not Very Important (1) to Very Important (7).

Procedures

Athletic administrators were informed of the purpose of the study prior to questionnaire distribution. A research assistant at each institution circulated the survey to all administrators and coaches. Nonresponse was handled by randomly selecting 10% ($n = 17$) of those individuals not completing the original questionnaire for direct telephone follow-up.

Analyses of the responses to the questionnaire showed no significant differences between respondents and nonrespondents ($F1,371) = 1.28$, $p < .27$). Multivariate Analysis of Variance (MANOVA), Repeated Measures Analysis of Variance, and Tukey's post hoc analyses were used to assess the relative differences in rating of performance indicators.

Results

The means and standard deviations of the six evaluation criteria dimensions for the administrator and coach subgroups, and the total data set are provided in Table 3. The numbers above the mean scores indicate the ranking of importance. Similar ratings by administrators and coaches are noted on all dimensions. Each dimension ranking by each subgroup indicates that it tested significantly different from the

Table 3 Means and Standard Deviations of Appraisal Criteria Dimensions by Subgroups

Criteria dimensions	Group		
	Combined ($n = 434$)	Administrators ($n = 77$)	Coaches ($n = 357$)
D3: Direct task behaviors	6.26^{1}	6.19^{1}	6.27^{1}
	± 0.58	± 0.55	± 0.59
D1: Team products	5.34^{2}	5.22^{3}	5.36^{2}
	± 0.90	± 0.67	0.94
D5: Administrative	5.21^{3}	5.63^{2}	5.15^{2}
maintenance behaviors	± 0.80	± 0.70	± 0.81
D4: Indirect task behaviors	4.96^{4}	5.35^{3}	4.87^{4}
	± 1.15	± 0.78	± 1.20
D6: Public relations	4.37^{5}	4.60^{5}	4.31^{5}
maintenance behaviors	± 1.03	± 0.96	± 1.03
D2: Personal products	3.69^{6}	3.82^{6}	3.66^{6}
	± 0.99	± 0.98	± 0.99

Note. Superscripts indicate dimension ranking of importance within groups, in order of decreasing importance as assessed by significant mean differences.

one above or below it or equal ranking if there was no difference. The direct task behaviors dimension was considered the most important dimension, whereas, the personal products dimension was rated the lowest.

MANOVAs were computed separately for the administrator (position × gender × experience) and coach data (sport coached × gender × experience × employment status x gender of team coached). The results showed minimal significant effects in all instances (see Table 4 & 5), and thus these independent variables are not further discussed within the scope of this paper. Next, the administrator and coach data were pooled, and a group (administrator vs. coach) × gender experience (1-4; 5-9; 10+ years) MANOVA was executed (see Table 6). Because the pooled data MANOVA analyses indicated that only group differences were significant, a 2 (director vs. coach) × 6 (performance evaluation criteria dimension) repeated measures ANOVA was used to assess the relative weighting of the six

Table 4 Summary of MANOVA Involving Dimensions of Performance Evaluation and Selected Independent Variables for Administrators

Source	Wilks' lambda	Exact *F*	*df*	Signif.
Position (A)	0.89	1.26	6,60	.291
Sex (B)	0.97	0.33	6,60	.920
Experience (C)	0.77	1.40	12,120	.173
A × B	0.96	0.47	6,60	.827
A × C	0.80	1.15	12,120	.330
B × C	0.85	0.85	12,120	.602
A × B × C	0.96	0.39	6,60	.884

Table 5 Summary of MANOVA Involving Dimensions of Performance Evaluation and Selected Independent Variables for Coaches

Source	Wilks' lambda	Exact *F*	*df*	Signif.
Sport type (A)	0.99	0.37	6,301	.897
Sex (B)	0.98	1.26	6,301	.275
Experience (C)	0.97	0.88	12,602	.570
Job status (D)	0.98	0.82	6,301	.552
Team sex (E)	0.95	1.20	12,602	.282
A × B	0.94	2.99	6,301	.008
A × C	0.97	0.80	12,602	.651
A × D	0.99	0.76	6,301	.599
A × E	0.95	1.41	12,602	.155
B × C	0.97	0.73	12,602	.719
B × D	0.99	0.70	6,301	.649
B × E	0.97	0.67	12,602	.780
C × D	0.94	1.59	12,602	.090
C × E	0.91	1.15	24,1051	.283
D × E	0.99	0.70	6,301	.649

(continued)

Table 5 *(continued)*

Source	Wilks' lambda	Exact *F*	*df*	Signif.
A × B × C	0.99	0.54	6,301	.781
A × B × D	0.99	0.42	6,301	.867
A × B × E	0.99	0.56	6,301	.760
A × C × D	0.94	1.46	12,602	.136
A × C × E	0.91	1.15	24,1051	.284
A × D × E	0.97	0.65	12,602	.798
B × C × D	0.98	0.62	12,602	.825
B × C × E	0.97	0.86	12,602	.588
B × D × E	0.98	0.80	6,301	.569
C × D × E	0.93	0.89	24,1051	.623
A × B × C × D	0.98	1.20	6,301	.306
A × B × C × E				
A × B × D × E	0.99	0.26	6,301	.954
A × C × D × E	0.97	1.40	6,301	.216
B × C × D × E	0.96	2.26	6,301	.038
A × B × C × D × E				

Table 6 Summary of MANOVA Involving Dimensions of Performance Evaluation and Selected Independent Variables for Administrators and Coaches

Source	Wilks' lambda	Exact *F*	*df*	Signif.
Position (A)	0.93	4.98	6,417	.000
Sex (B)	0.99	0.53	6,417	.786
Experience (C)	0.95	1.68	12,834	.076
A × B	0.99	0.62	6,417	.714
A × C	0.97	0.96	12,834	.488
B × C	0.98	0.80	12,834	.646
A × B × C	0.98	0.85	12,834	.596

Table 7 Summary of Repeated Measures ANOVA Involving Dimensions of Performance Evaluation by Position

Variable	*df*	SS	MS	*F*
Position (A)	1	14.06	14.06	5.05*
Error	432	1201.53	2.78	
Evalu. dimen. (B)	5	905.33	181.07	391.66***
A × B	5	19.25	3.85	8.33***
Error	2160	998.57	0.47	

*$p < .05$; ***$p < .001$.

dimensions within the coach and administrator group (see Table 7). These analyses were followed by Tukey's post hoc analyses as warranted to identify the specific subgroup differences.

Ratings of Dimensions by Administrators and Coaches

Results of the MANOVA showed only the main effect of group (administrator vs. coach) was significant ($F_{6,417} = 4.98, p < .001$). Univariate analyses showed that the main effect of group on team products ($F_{1,422} = 4.67, p < .05$) and direct task behaviors ($F_{1,422} = 5.85, p < .05$) was significant, and that the main effect of experience on direct task behaviors ($F_{2,422} = 4.88, p < .01$) and administrative maintenance behaviors ($F_{2,422} = 3.89, p < .05$) was also significant. Post hoc procedures did not uncover significant mean differences on direct task behaviors, but mean differences did occur on administrative maintenance behaviors between most experienced (10+ years, $m = 5.07$), moderately experienced (5-9 years, $m = 5.17$), and least experienced (1-4 years, $m = 5.40$) administrators and coaches. In summary, coaches rated team products and direct task behaviors significantly higher than administrators did, and the least experienced administrators and coaches tended to rate administrative maintenance behaviors as more important than did colleagues with more experience.

Rankings of Dimensions by Administrators and Coaches

The foregoing analyses were concerned with subgroup differences pertaining to each of the six dimensions. To verify the relative rating of the six dimensions, a repeated measures ANOVA was applied to assess the rankings. Because previous analyses showed minimal differences in the ratings of the individual dimensions due to gender and experience, these two variables were eliminated. Due to substantial differences between administrators and coaches, the dichotomy of administrator-coach was retained as the grouping variable.

The two levels of group (administrator and coach) by six dimensions repeated measures ANOVA showed that main effects of group ($F_{1,432} = 5.05, p < .05$) and dimensions ($F_{5,2160} = 391.66, p < .001$) and their interaction ($F_{5,2160} = 8.33, p < .001$) were significant. Tukey's post hoc analyses revealed that within both the administrator and coach group each of the means was found to differ significantly across the six criteria dimensions except team products and indirect task behaviors for administrators, and team products and administrative maintenance behaviors for coaches. Overall, the significant differences found between the mean ratings of the dimensions supports the six dimensional substructure.

The overall rankings attached to each dimension are clearly similar, which shows that the two groups perceive the importance of the performance criteria similarly (Table 5), although the order of importance differed slightly between the two groups.

Additionally, the repeated measures ANOVA results for the total data set over the six dimensions were significant ($F_{5,2165} = 710.35, p < .001$). Post hoc tests showed that each of the dimension mean ratings were significantly different from one another. The administrators and coaches en mass rated direct task behaviors as the most important criteria dimension ($m = 6.26$), followed by team products ($m = 5.34$), administrative maintenance behaviors ($m = 5.21$), indirect task behaviors ($m = 4.96$), public relations behaviors ($m = 4.37$) and personal products ($m = 3.69$).

Discussion

A remarkable finding of this study was the depth of correspondence that administrators and coaches gave to the evaluation criteria dimensions. Their mean ratings of importance for each dimension are displayed in Figure 4.

For administrators, coaches, and all their subgroups, direct task behaviors, a category of behavioral process factors, was commonly perceived to be the most important. This dimension encompasses those coaching evaluation criteria that evaluate the specific abilities or skills used directly in the day-to-day application of coaching. The literature pertaining to evaluation criteria strongly affirms the

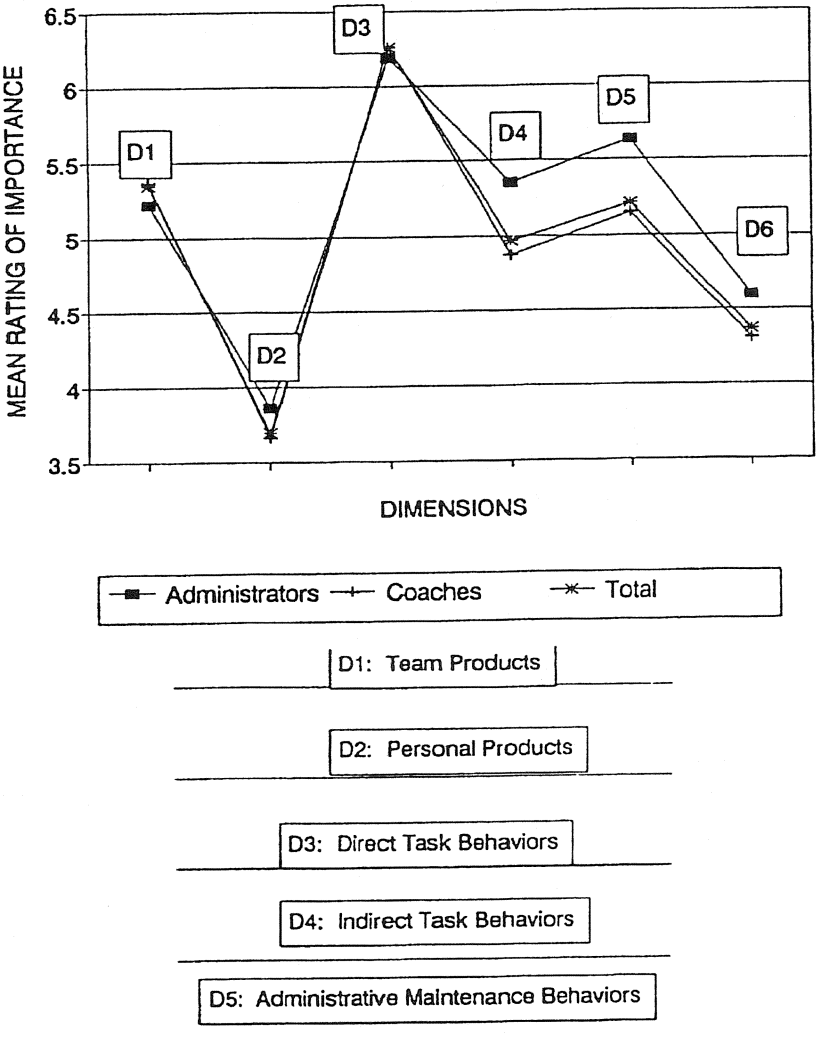

Figure 4 — Mean rating of importance attached to dimensions of evaluation criteria.

importance of using job-specific evaluation criteria (Murphy & Cleveland, 1991). This study empirically confirmed this notion for college coaching.

Coaches and administrators differed in their rating of the second most important dimension. Administrators rated administrative maintenance behaviors second in importance, whereas coaches rated both team products and administrative maintenance behaviors second in criteria dimension importance. It is possible that each group identifies with their own particular job orientation. Administrative maintenance behaviors include the management activities required to conduct the business aspects of sport. Administrators rated the criteria within this dimension as second most important, perhaps because their job orientation as managers results in a focus on management activities, or because the coach's ability to manage his or her sport program directly affects the job environment of the administrator. In parallel, coaches rated the outcomes of coaching that accrue only to the team or individual athletes (team products) second in importance. Perhaps this is because outcomes are public indicators of the coach's performance and most in line with the coach's focus. One factor often included in the coach's job focus is goal achievement for athletes (Cratty, 1983), which may account for higher ratings on items related to achieving goals (i.e., product factors). Also, it should be noted that although mean differences between administrators' and coaches' ratings of important criteria were statistically significant, they were also very small (5.22 vs. 5.36, respectively).

Of interest is that the criteria dimensions rated the highest by administrators were both behavioral process factors, task-related process factors (D3) and maintenance-related process factors (D5). It is apparent that administrators consider factors related to the process of coaching more important than the products. On the other hand, coaches rated one dimension from each of the main theoretical divisions as having the most significant evaluation criteria. Behavioral process factors (D3: direct task behaviors) and behavioral product factors (Dl: team products) gained their most favor, suggesting that coaches view evaluation on specific coaching abilities and the outcomes of those abilities (manifested through athlete/team performance) as the major criteria for evaluating their performance.

Both administrators and coaches were compatible in rating public relations behaviors and personal products, respectively, as the least important criteria dimensions. This suggests lesser importance attributed to liaison activities to the sport community and to outcome factors that mostly benefit the coach. Perhaps liaison activities, while important, are viewed as adjuncts to the coaching role and that personal recognition is not always attributed to coaching effectiveness. For example, evaluating a coach on frequency of receiving coaching excellence awards might include many factors of a political (criteria of bestowing the awards) and environmental (number of awards presented) nature, which in turn could lead to an unfair and ineffective evaluation.

The assessment of grouped data for administrators and coaches on the appraisal criteria dimensions revealed consistent ratings within and between the two groups. There were no significant differences among the administrator groupings on rating the importance of the six dimensions. Athletic directors, coordinators, and administrative assistants held comparable views about which criteria were most important in evaluating coaches. This congruence of perceptions held across gender and experience levels. Because of the high representation of Canadian

administrators, this finding lends credence to the strength of these criteria, which in turn, is a contribution, however modest, to the management of Canadian university sport.

The coaches' ratings of the appraisal criteria dimensions showed congruency across independent factors of gender and experience levels. There were no significant differences in their ratings. The demonstrated congruency in rating the dimensions of evaluation criteria among this population can be compared to studies investigating rater characteristics. Many authors (Elmore & Lapointe, 1975; Rose, 1978; Schmitt & Lappin, 1980) report little or no effect on performance ratings resulting from rater gender, a similar result to those found in this study. Early studies reported job experience as a factor positively affecting the quality of ratings, although the reason remained unclear (Cascio & Valenzi, 1977; Klores, 1966; Mandell, 1956). Additionally, studies to assess the effects of the type of rater did suggest that different raters (self, subordinate, supervisor) are likely to have different perspectives on job performance (Centra, 1975; Kraut, 1975; Lawler, 1967). However, in this study, the perception of important evaluation criteria did not differ between coaches and administrators. The congruence seen is possibly specific to the population involved.

In addition to assessing the importance attached to the criteria dimensions, an attempt was made to assess whether this population supported the framework of criteria dimensions proposed in the theoretical model. A summary of the importance attached to the criteria dimensions by administrators and coaches was previously presented in Table 1. With the exception of team products and indirect task behaviors for administrators, and team products and administrative maintenance behaviors for coaches, the means of each criteria dimension were significantly different from one another, lending support to the theoretical model. In addition, both behavioral product and behavioral process factors were perceived to contain important coaching evaluation criteria, a factor that further supports the theoretical model proposed in this study.

When dimensional importance was assessed for the total group of subjects, the results presented above were reaffirmed. Each of the mean ratings were different, lending support to the theoretical model. Direct task behaviors, team products, and administrative maintenance behaviors were respectively the top-ranked dimensions. It is interesting that these three dimensions are representative of the three main categories of criteria espoused by the theoretical model. This finding also lends support to the model.

While the affirmation of the importance of using different types of coaching performance criteria is a critical finding of this study, the harmonious support for behavioral process factors is also of great importance and perhaps worthy of a closer look. Murphy and Cleveland (1991) strongly suggest that evaluation criteria be job-specific, representative of factors most critical to performance effectiveness, and measurable. The theoretical criteria listed in this study are certainly job-specific and perceived to be critical to performance effectiveness by the populations assessed; however, no attempt has been made to investigate the measurability of the criteria. It is possible that the issue of measurement may be factored into the perceptions of important criteria, such that the criteria dimensions rated highest are also the most quantifiable. This may be an issue most specific to administrators who currently perform evaluations. Subsequently, the ease of measurement issue is worthy of further consideration and study.

Conclusion

Direct task behaviors were rated the most important evaluation criteria classification in support of the need for job-specific coaching performance criteria. In general, administrators and coaches, regardless of subpopulations within each group, held corresponding beliefs about criteria important for evaluating coaching performance. The integrity of the model's classification of performance evaluation criteria was verified by those who would use the model (administrators) and for those on whom the model would be used (coaches). This finding provides strong support for the theoretical model, a significant result toward establishing baseline evaluation criteria for Canadian college coaches.

References

Anderson, E.W. (1985). The faculty-coach. *Journal of Physical Education, Recreation, & Dance,* **56**(9), 16-18.

Astin, A. (1964). Criterion-centered research. *Educational and Psychological Measurement,* **24**(4), 807-822.

Barber, R., & Skoglund, F. (1981). Improving coaching through evaluation. *The Athletic Director,* **11**(3), 1,4.

Beer, M. (1987). Performance appraisal. In J.W. Lorsch (Ed.), *Handbook of Organizational Behavior* (pp. 286-300). Englewood Cliffs, NJ: Prentice-Hall.

Bennice, D.A. (1990, October). Evaluating division III professional staffs. *Athletic Administration,* **25,** pp. 16-18.

Buford, J.A., Burkhalter, B.B., & Jacobs, G.T. (1988). Link job descriptions to performance appraisals. *Personnel Journal,* **67**(6), 132-140.

Carroll, S.J., & Schneier, C.E. (1982). *Performance appraisal and review systems: The identification, measurement, and development of performance in organizations.* Glenview, IL: Scott, Foresman.

Cascio, W.F., & Valenzi, E.R. (1977). Behaviorally anchored rating scales: Effects of education and job experience of raters and ratees. *Journal of Applied Psychology,* **62,** 278-282.

Centra, J.A. (1975). *The influence of different directions on student ratings of instruction* (ETS RB75-28). Princeton, NJ: Educational Testing Service.

Cratty, B.J. (1983). *Psychology in Contemporary Sport.* Englewood Cliffs, NJ: Prentice-Hall.

Cummings, L.L., & Schwab, D.P. (1973). *Performance in organizations: Determinants and appraisal.* Glenview, IL: Scott, Foresman.

Darling-Hammond, L. (1986). A proposal for evaluation in the teaching profession. *The Elementary School Journal,* **86**(4), 1-21.

Darling-Hammond, L., Wise, A.E., & Pease, S.R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research,* **53**(3), 285-328.

Deets, N.R., & Tyler, D.T. (1986). How Xerox improved its performance appraisals. *Personnel Journal,* **65**(4), 50-52.

Elmore, P.B., & LaPointe, K.A. (1975). Effect of teacher sex, student sex, and teacher warmth on the evaluation of college instructors. *Journal of Educational Psychology,* 67, 368-374

Fleishman, E., & Quaintance, M. (1984). *Taxonomies of human performance: The description of human tasks.* New York: Academic Press.

Ilgen, D.R., & Barnes-Farrell, J.L. (1984). Performance planning and evaluation. In J.E. Rosenzweic & F.E. Kast (Eds.), *Modules in Management.* Chicago: Science Research Associates.

James, L. (1973). Criterion models and construct validity for criteria. *Psychological Bulletin,* **80**(1), 75-83.

Klores, M.S. (1966). Rater bias in forced-distribution ratings. *Personnel Psychology,* **19**(4), 411-421.

Knapp, M.S. (1982). *Toward the study of teacher evaluation as an organizational process: A review of current research and practice.* Menlo Park, CA: Educational and Human Services Research Center, SRI International.

Knorr, J. (1989). Evaluating athletic personnel. *Texas Coach,* **34**(3), 20-21.

Kraut, A.I. (1975). Prediction of managerial success by peer and training-staff ratings. *Journal of Applied Psychology,* **60**(1), 14-19.

Kurtz, R.S., Mueller, J.J., Gibbons, J.L., & Dicataldo, F. (1988). Faculty performance: Suggestions for the refinement of the concept and its measurement. *Journal of Higher Education,* **60**(l), 43-58.

Landy, F.J., & Farr, J.L. (1983). *The measurement of work performance: Methods, theory, and applications.* New York: Academic Press.

Lane, B.A. (1990). Personnel evaluation: From problems to school improvement. *Journal of Research and Development in Education,* **23**(4), 243-249.

Lawler, E.E., III (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology,* **51,** 369-381.

Lawler, E.E., Mohrman, A.M., & Resnick, S.M. (1984). Performance appraisal revisited. *Organizational Dynamics,* **13**(l), 20-35.

Leland, T. (1988, Nov.-Dec.). Evaluating coaches—Formalizing the process. *Journal of Physical Education, Recreation & Dance,* **59**(9), 21-23.

MacLean, J.C. (1993). Coaching evaluation: A guide for establishing job-specific criteria. *Applied Research in Coaching and Athletics Annual—1993,* pp. 44-60.

MacLean, J.C. (1994). Personnel evaluation in education: A model for job-specific criteria. *The Clearing House,* **68**(l), 47-51.

MacLean, J.C., & Chelladurai, P. (1995). Dimensions of coaching performance: Development of a scale. *Journal of Sport Management,* **9**(2), 194-207.

MacLean, J.C., & Zakrajsek, D. (1994). Evaluating athletic coaches: A descriptive analysis of Canadian universities. *CAHPER Journal,* **60**(2), 5-10.

Mandell, M.M. (1956). Supervisory characteristics and ratings: A summary of recent research. *Personnel Psychology,* **32,** 435-440.

Margolis, J.A. (1979). A better way to evaluate your coaches. *Athletic Administration,* **14**(2), 12-13.

Martin, T.P., Arena, L.L., Rosencrans, R.E., Hunter, L.E., & Holly, P.W. (1986, August). College level coaching: An evaluation. *Journal of Physical Education, Recreation & Dance,* **57**(6), 49-52.

Murphy, K.R., & Cleveland, J.N. (1991). *Performance appraisal: An organizational perspective.* Boston: Allyn and Bacon.

Norcross, T.H. (1986, July 18-23). Performance appraisal for coaches. In Bernard Wright (Chair), *Coach education: Preparation for a profession.* Proceedings of the VIII Commonwealth and International Conference on Sport, Physical Education, Dance, Recreation and Health, Glasgow, Scotland.

Peterson, P.L., & Comeaux, M.A. (1990). Evaluating the systems: Teachers' perspectives on teacher evaluation. *Educational Evaluation and Policy Analysis,* **12**(l), 3-24.

Rieder, G.A. (1973, July/August). Performance review—A mixed bag. *Harvard Business Review, 61-67.*

Rose, G.L. (1978). Sex effects on effort attributions in managerial performance evaluation. *Organizational Behavior and Human Behavior,* **21,** 367-378.

Sabock, R.J. (1985). *The coach.* Champaign, IL: Human Kinetics.

Salvendy, G., & Seymour, W. (1973). *Prediction and development of industrial work performance.* New York: Wiley.

Schermerhorn, J.R., Hunt, J.G., & Osborn, R.N. (1985). *Managing organizational behavior,* 2nd Ed. New York: Wiley.

Schmitt, N., & Lappin, M. (1980). Race and sex as determinants of the mean and variance of performance ratings. *Journal of Applied Psychology,* **65,** 428-435.

Smith, P.C. (1976). Behaviors, results, and organizational effectiveness. In M. Dunnette *(Ed.), Handbook of Industrial and Organizational Psychology* (pp. 745-775). Chicago: Rand-McNally.

Torrington, D., Weightman, J., & Johns, K. (1989). *Effective management—People and organization.* Englewood Cliffs, NJ: Prentice-Hall.

Vecchio, R.P. (1988). *Organizational behavior.* Chicago: Dryden Press.

Wexley, K.N., & Latham, G.P. (1981). *Developing and Training Human Resources in Organizations.* Glenview, IL: Scott, Foresman.