

## INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.
2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University  
Microfilms  
International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106  
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

7913052

**KNOPPERS, ANNELIES**

**A BEHAVIORALLY BASED APPRAISAL OF COACHING  
PERFORMANCE IN WOMEN'S ATHLETICS AT MICHIGAN  
STATE UNIVERSITY.**

**THE UNIVERSITY OF NORTH CAROLINA AT  
GREENSBORO, ED.D., 1978**

University  
Microfilms  
International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

© 1979

ANNELIES KNOPPERS

ALL RIGHTS RESERVED

A BEHAVIORALLY BASED APPRAISAL OF COACHING  
PERFORMANCE IN WOMEN'S ATHLETICS  
AT MICHIGAN STATE UNIVERSITY

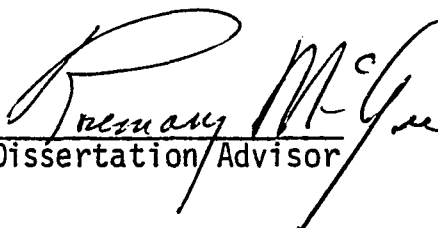
by

Annelies Knoppers

A Dissertation Submitted to  
the Faculty of the Graduate School at  
The University of North Carolina at Greensboro  
in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Education

Greensboro  
1978

Approved by

  
Dissertation Advisor

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of the Graduate School at the University of North Carolina at Greensboro.

Dissertation  
Adviser

Joan Mary Mc Gee

Committee Members

William A. Powers

Margaret A. Marday

Sarah M. Robinson

Pearl Berlin

September 18, 1978  
Date of Acceptance by Committee

KNOPPERS, ANNELIES. A Behaviorally Based Appraisal of Coaching Performance in Women's Athletics at Michigan State University. (1978).  
Directed by: Dr. Rosemary McGee. Pp. 173.

The purpose of this study was to construct a multitrait multi-rater model for the appraisal of the 10 head coaches in the women's athletic program at Michigan State University in 1977-78. The resulting instrument was a behaviorally anchored rating scale (KBARS) consisting of 10 dimensions.

The construction of the KBARS involved members of the rater population. They submitted 623 examples of effective and ineffective coaching behaviors. These incidents were refined to 299 statements. On the basis of the literature pertaining to coaching and the content of these incidents, 12 categories or dimensions of coaching performance were defined. A stratified sample of 43 raters-to-be allocated each statement to a dimension and assigned the item a numerical value from one to seven. In order to be included in the final KBARS, items and dimensions had to meet modal assignment, variability, and agreement criteria.

Ten dimensions and 123 items met all the criteria for inclusion. The resulting dimensions were: (1) Scheduling; (2) Practice Content; (3) Team Selection; (4) Personality; (5) Team Discipline; (6) Psychological Support; (7) Communication; (8) Professionalism; (9) Knowledge; and, (10) Public Relations/Recruiting. The statements became the behavioral anchors for their respective dimensions.

The KBARS was completed by a team, its head coach, assistant coach, trainer, peer, and women's athletic director at the end of a sports season. A total of 173 ratings were completed which included 128 varsity athletes and 45 staff members. The raters rated a coach on a seven-point scale on each of the 10 dimensions using the behavioral anchors as a guide and commented on and/or gave examples of behaviors by the coach which substantiated the ratings.

The data were analyzed by a principal component factor analysis procedure. This yielded one distinct factor, the KBARS Factor, indicating that the dimensions were highly related and that the instrument functioned as a unity so that ratings could be summed across dimensions. The intercorrelations between dimensions ranged from .31 to .72. The Personality dimension overlapped the most with the other dimensions while the Public Relations/Recruiting dimension overlapped the least.

A psychometric analysis of the KBARS data revealed response set bias. A high degree of leniency, especially in the self and peer ratings, and a moderate halo effect were shown. The extent to which these were errors was unclear since the comments by the raters indicated that the coaches had performed effectively.

The extent of the internal reliability was assessed in two ways. The scale reliability coefficients ranged from .83 to 1.00 depicting the extent of agreement among those in the sample on the values assigned to the behavioral anchors. The intraclass coefficients ranged from .61 to .87 indicating the extent of between-rater

agreement on dimensions. Since current methods of obtaining external reliability coefficients are inappropriate for use on such a scale, external reliability was not determined.

The developmental procedure to construct the KBARS ensured a high degree of content validity. There was a significant difference between the ratings of the athletes and the staff indicating a low degree of convergent validity. The degree of discriminant validity was moderate based on the size of the inter-dimension correlation coefficients. On the whole, the degree of validity and reliability shown by the KBARS ratings were judged to be adequate ensuring that the KBARS measured coaching performance in a valid and reliable manner.

## ACKNOWLEDGMENTS

The completion of this dissertation and doctoral work reflects the contributions of many. To the following contributors I express deep appreciation and the acknowledgment that the extent of their help and influence was immeasurable:

My parents who have always stressed development of God-given talents and then supplied the example, love, and support enabling me to make this dream a reality.

Dr. Rosemary McGee who challenged and encouraged me and in so doing was an ideal role model.

The members of my doctoral committee who gave many helpful suggestions and assistance.

Dr. Nell Jackson who inspired this study and throughout its duration gave sage advice, moral support, and infinite encouragement.

Nancy Van Noord and Bartheke Knoppers, two "sisters," who warmly and loyally supported me in many varied ways ensuring that I grew not only as a researcher but also as a person.

The KBARS is dedicated to assistant coach, Nancy Steel, and the members of the volleyball team at Michigan State University in the hope that the results from instruments such as these will assist their quest for athletic excellence and will enhance coach-athlete relationships.



## TABLE OF CONTENTS

	Page
APPROVAL PAGE . . . . .	ii
ACKNOWLEDGMENTS . . . . .	iii
LIST OF TABLES . . . . .	vii
CHAPTER	
I. INTRODUCTION . . . . .	1
Statement of the Problem . . . . .	5
Significance of the Study . . . . .	5
Assumptions Underlying the Study . . . . .	7
Scope of the Study . . . . .	8
Definition of Terms . . . . .	8
Michigan State University and Its Athletic Program . . . . .	10
Basketball . . . . .	13
Cross Country . . . . .	13
Field Hockey . . . . .	13
Golf . . . . .	14
Gymnastics . . . . .	14
Softball . . . . .	15
Swimming and Diving . . . . .	15
Tennis . . . . .	16
Track and Field . . . . .	16
Volleyball . . . . .	17
II. REVIEW OF RELATED LITERATURE AND RESEARCH . . . . .	18
Traditional Approaches to the Evaluation of Coaches . . . . .	18
Personality Profiles of Coaches . . . . .	19
Athletes' Perceptions of Coaches . . . . .	22
Beliefs Concerning Essential Coaching Attributes . . . . .	30
Current Methods for Evaluating Coaching Performance . . . . .	32
Summary . . . . .	34
Development and Use of Behaviorally Anchored Rating Scales (BARS) . . . . .	35
Development of BARS . . . . .	35
The BARS in Comparison with Other Scales . . . . .	39
Summary of Studies Involving BARS . . . . .	43

CHAPTER	Page
III. PROCEDURES . . . . .	44
Phase I: Item and Dimension Development . . . . .	45
Orientation of the Rater Population . . . . .	45
Collection of Behavioral Statements . . . . .	46
Refinement of Statements . . . . .	47
Dimension Selection and Refinement . . . . .	47
Phase II: Retranslation Procedure . . . . .	48
Selection of Sample . . . . .	48
Retranslation by the Sample . . . . .	49
Phase III: Analysis of Retranslation Results . . . . .	51
Retention of Dimensions . . . . .	51
Retention of Items . . . . .	51
Between Group Agreement . . . . .	52
Phase IV: Construction of the KBARS . . . . .	53
Final Selection of Items and Dimensions . . . . .	54
Appraisal Booklet . . . . .	55
Phase V: Administration of the KBARS . . . . .	56
Phase VI: Statistical Analysis . . . . .	57
Phase VII: Psychometric Analysis of the KBARS . . . . .	58
Response Set Bias . . . . .	58
Reliability . . . . .	59
Validity . . . . .	60
IV. FINDINGS AND DISCUSSION . . . . .	62
Phase I: Item and Dimension Development . . . . .	62
Collection of Behavioral Statements . . . . .	63
Phase II: Retranslation Procedure . . . . .	66
Selection of Sample . . . . .	66
Retranslation by the Sample . . . . .	66
Phase III: Analysis of Retranslation Results . . . . .	66
Retention of Dimensions . . . . .	68
Retention of Items . . . . .	70
Between Group Agreement . . . . .	72
Dimensions and Items Surviving All Criteria . . . . .	74
Scale Reliability . . . . .	76
Phase IV: Construction of the KBARS . . . . .	78
Phase V: Administration of the KBARS . . . . .	79
Ratings by the Population . . . . .	79
Discussion . . . . .	81
Phase VI: Statistical Analysis . . . . .	82
Correlation Matrix . . . . .	82
Factor Analysis . . . . .	82
Discussion . . . . .	85
Phase VII: Psychometric Analysis . . . . .	87
Response Set Bias . . . . .	87
Reliability . . . . .	92
Validity . . . . .	94
Part III of the KBARS . . . . .	99
Summary . . . . .	101

CHAPTER	Page
V. SUMMARY AND CONCLUSIONS . . . . .	103
Summary . . . . .	103
Conclusions . . . . .	106
Feasibility of BARS for Appraising Coaching Performance . . . . .	106
Implications of the KBARS Results for Coaching . . . . .	110
Conclusions . . . . .	112
Recommendations for the KBARS . . . . .	113
Suggestions for Further Study . . . . .	116
BIBLIOGRAPHY . . . . .	117
APPENDIX A	
Letter to Athletes . . . . .	127
Letter to Coaches . . . . .	
Behavioral Statements - Phase I . . . . .	
Examples of Behavioral Statements . . . . .	
APPENDIX B	
Coaches Evaluation Project - Phase II . . . . .	136
APPENDIX C	
Coaching Evaluation Packet (KBARS) . . . . .	158

## LIST OF TABLES

Table	Page
1. Retranslation Sample . . . . .	67
2. Number of Items Surviving Each Criterion . . . . .	69
3. Dimension Statistics for Retained Items . . . . .	71
4. Size of Population Using KBARS . . . . .	80
5. Correlation Matrix of KBARS Dimensions . . . . .	83
6. Factor Analysis of KBARS Dimensions . . . . .	84
7. Summary Data of KBARS Dimensions by Rater Groups . . . . .	88
8. Medians and Ranges of Interdimension Correlations of KBARS . . . . .	89
9. Reliability Coefficients for KBARS . . . . .	93
10. Analysis of Variance of Groups by Factor Scores . . . . .	96
11. Classification of Comments/Incidents of Part III . . . . .	100

## CHAPTER I

### INTRODUCTION

The number of coaches involved in athletics has increased greatly during the past few years due to the growth of athletic programs for girls and women. This increase has necessitated a focus on issues dealing with the various aspects of coaching roles and performance. One such issue particularly appropriate to institutions of higher education deals with the performance appraisal of coaches.

Since evaluation is inevitable in higher education (Doyle, 1975; Dressel, 1961) coaches in educational institutions must also be evaluated. Such an appraisal should maintain or improve the quality of coaching by exposing inadequacies as well as strengths. Coaching performance should show evidence of a close approximation to some ideal. Intercollegiate athletics are part of higher education. Institutions of higher education are "dedicated to the advancement of knowledge, to the development of reasoning and judgment in its students" (Dressel, 1961, p. 25). Thus, evaluations carried out under the auspices of such institutions should have a sound basis and be rigorously developed. This applies to the appraisal of coaching performance in intercollegiate athletics as it does to other assessments in the institutions.

In the past, the majority of coaches at the intercollegiate level have been part of the men's athletics program. Their success

was and is determined by one criterion, mainly the win/loss record (Sage, 1975). If the production of winning teams is to be the major standard for coaching success in women's intercollegiate athletics, then such a record should be the main criterion in the performance appraisal of its coaches. However, individuals involved in women's athletics assert that the athletic program for women should provide a new model for intercollegiate competition (Hunt, 1974; Ley, 1973; Poindexter, 1973). Such a model should be consonant with the educational philosophy of the institution rather than being based on winning and on revenues. Part of this model should encompass educationally sound procedures to appraise coaching performance. The development of such an evaluation system specific to an institution's women's athletic program is the main purpose of this study.

The assessment of college and university faculty members has been sporadic and has just recently become the focus of intensive research (Doyle, 1975; Eble, 1970). Rating scales have been the primary measurement tool for such assessments. Faculty members have been rated by students, peers, administrators, and themselves. The results of investigations dealing with the data generated by these scales have been equivocal and often contradictory. In industry where ratings are also used a great deal, the same problems abound (Dunnette, 1966, 1976; Schneider, 1976). Difficulties have been due primarily to the inadequacies and ambiguities of the rating instruments. The major deficiencies in most rating scales stem from an inadequate sampling of the domain of relevant job behaviors by the

scales' developers and the formation of response sets by the scales' users (Dunnette, 1976; Jordan, 1976).

Smith and Kendall (1963) attempted to minimize the problems associated with rating scales by developing a behaviorally anchored rating scale (BARS) for evaluating nurses. Since its creation, BARS has been used in the performance appraisal of a variety of occupations such as fire-fighters (Dickinson and Tice, 1973), college instructors (Harari and Zedeck, 1974), military personnel (Borman and Dunnette, 1975), grocery clerks (Fogli, Hulin and Blood, 1971) and engineers (Landy and Guion, 1970). BARS consists of scales which are anchored by statements descriptive of actual job behaviors. Members of the rating population are involved in scale construction through their contributions of effective and ineffective incident statements, their definitions of essential qualities or dimensions of the job, and their assignment of values to the incidents and of incidents to dimensions. This is known as "retranslation." The scales for each dimension are then stated in expectation terminology to obviate the sampling problem. An example of one dimension of such a scale is given in Appendix C.

The BARS method has theoretical advantages over traditional rating methods (Smith and Kendall, 1963; Bernardin, LaShells, Smith and Alvares, 1976; Jordan, 1976; Schwab, Heneman and DeCotiis, 1975). The scales are rooted in actual observable behaviors, are constructed by persons similar to those who use the scales, and consist of very specific and nonambiguous qualities and incidents.

Such advantages offer the promise of yielding an increase in the reliability of the ratings as well as a decrease in response set errors (Dunnette, 1976; Zedeck and Blood, 1974). The BARS procedure has been selected as the basis for the development of a model for appraising coaching performance in the present study.

Although the visible part of coaching takes place on the courts and fields, less obvious facets are part of the job. According to Resick and Erickson (1975), Sage (1975), and Deatherage and Reid (1977) the coach also has responsibilities for scheduling, taking care of equipment, establishing (and working within) a budget, building positive public relations, coordinating a staff, arranging transportation, recruiting and retaining players, as well as contributing to the overall athletic program. The extent of the exact degree of coaching duties varies from institution to institution. The literature pertaining to coaching indicates that certain qualities are essential to success (Hendry, 1969; LaGrand, 1970; Percival, 1971; Clark, 1975). The actual importance of these qualities with respect to coaching performance has not been rigorously validated. Administrators, athletes, other coaches, support personnel and the coach herself attach varied degrees of significance to these qualities and responsibilities. It is fitting, therefore, that the model for appraising coaching performance utilize the multitrait, multirater method. It not only must include an adequate sampling of coaching behaviors but must also give guidelines for determining



qualities which are essential to job success and determining rater groups who are most accurate in judging these dimensions.

#### Statement of the Problem

The purpose of this study is to develop a multitrait multi-rater behaviorally anchored appraisal model of coaching performance of ten head coaches of the women's athletic program at Michigan State University. More specifically, the following subproblems will be studied:

1. What common elements of coaching performance observed in ten different sports are generated by the use of a behaviorally anchored rating scale (KBARS)?
2. To what extent do KBARS procedures yield a valid and reliable means of appraising coaching performance?

#### Significance of the Study

The lack of accuracy of performance ratings has been a problem in both industry and education (Doyle, 1975; Dunnette, 1976). The factor analysis of the results of such ratings consequently have been vague, difficult to interpret and not always replicable (Harari and Zedeck, 1974). Persons involved in coaching have long espoused the fact that coaches should be evaluated on criteria other than the win/loss record (Deatherage and Reid, 1977; Frost, 1971; Gallon, 1974; Ley, 1973; Resick and Erickson, 1975). Consequently, an appraisal system is needed which minimizes the problems cited above.

The use of a behaviorally anchored rating scale (BARS) may help to assuage this need.

A BARS focuses attention on actual behaviors rather than on global overall effects (Schwab, Heneman and DeCotiis, 1975; Smith and Kendall, 1963). Involvement by the raters in the scale's construction contributes to the meaningfulness of the instrument since the behaviors are stated in "job" language which lessens the extent of ambiguity which is prevalent in most rating scales (Dickinson and Tice, 1973; Millard, 1975). Since the raters are dealing with familiar behaviors there should be a decrease in response errors, an increase in factor clarity and an increase in inter-rater reliability (Dunnette, 1976; Smith, 1976).

This is the first time that a BARS has been applied to the performance appraisal of coaches. Its usefulness as well as the multitrait multirater perspective should yield results highly amenable to counselling and feedback. Thus the instrument has the potential to fulfill its purpose of improving the quality of coaching performance. The performance dimensions and items on the rating scale can yield a basic set of concepts and measures which could be directly applied as guidelines for suitable coaching behaviors, as clarification of coaching responsibilities and job analyses, and, at the same time, provide insight into the relationship between the espoused philosophy of athletics at an institution and the conduct of its program.

By the very nature of the steps involved in its development the resultant appraisal model will be specific to the women's athletic program at Michigan State University. If the use of this model is found feasible and has merit, then the procedures can be replicated at other institutions.

In addition, it is hoped that such an endeavor will stimulate research in coaching behaviors and their evaluation. If it is true that performance is geared to expectation, then the development and use of such a model can help to shape the direction and perspective of those involved in women's athletics by clarifying the value system currently in operation.

#### Assumptions Underlying the Study

Basic to the research are the following assumptions:

1. Coaching behaviors which are essential to successful performance can be described by coaches, athletes, support personnel and administrators.
2. The dimensions of the KBARS adequately represent the universe of qualities essential to coaching performance.
3. The scales used to evaluate each dimension adequately sample the full range of behaviors related to that dimension.

4. The BARS procedures yield unidimensional scales so that a dimension continuum ranges from effective to ineffective behaviors.
5. The selected rating groups are homogeneous.

### Scope of the Study

The following are boundaries of the present study:

1. Raters include the female athletes, coaches, support personnel, as well as the Assistant Athletic Director for Women at Michigan State University, East Lansing, Michigan.
2. The development and scope of the KBARS is limited to the input of the rater population.
3. Data from the model are collected during the 1977-78 academic year.

### Definition of Terms

The following terms are defined as they are used in this study:

Athlete: A female student at Michigan State University who was listed on the eligibility list for 1977-1978 of the Association for Intercollegiate Athletics for Women (AIAW) and is listed in her team's scorebook for two-thirds of a team's matches or games.

Behaviorally Anchored Rating Scale (BARS): An appraisal procedure which assesses performance in multidimensional and behavior specific terms. Its construction involves the potential raters.

Coaching Behavior: Behavior engaged in by the coach which is related to the performance of his/her duties and responsibilities as head coach. This excludes those behaviors and activities carried out in conjunction with other departments at Michigan State University.

Critical Incidents: Episodes which are selected by the rater population as being illustrative of ineffective and effective coaching behaviors.

Dimension: A homogeneous category representing a quality considered important for performance (Bernardin, Alvares and Cranny, 1976).

Effective Job Behavior: Behavior engaged in by the coach which is perceived to facilitate the success of her/his sport's program (Jordan, 1976).

Head Coach: The person designated by the Assistant Athletic Director-Women to be in charge of one of the women's intercollegiate sport programs at Michigan State University for 1977-1978.

Ineffective Job Behavior: Behavior engaged in by the head coach which is perceived to inhibit the success of her/his sport's program (Jordan, 1976).

Junior Varsity Team Member: A player who competes as a member of this team for 75% of this team's matches or meets.

Knoppers' Behaviorally Anchored Rating Scale (KBARS): The BARS which was constructed for this study.

Rater Population: The total of all the athletes, coaches, support personnel and administrators involved in the women's athletic program at Michigan State University 1977-1978.

Rater Subgroups or Subsets: The division of the rater population into athletes, head coaches, support personnel, and administrator.

Sample for KBARS: A stratified sample representative of the rater population which was involved in the retranslation aspect of the development of KBARS (Phase II).

Support Personnel: Persons who have been assigned one or more of the following functions in the women's athletic program: assistant coach, trainer, manager, statistician or intern.

Varsity Team Member: A player who is listed on this team's roster for 75% of the team's scheduled events.

Women's Athletics: The athletic program at Michigan State University which sponsors the following sports for women athletes: cross country, field hockey, swimming, diving, gymnastics, golf, tennis, softball, basketball, track and field, and volleyball.

#### Michigan State University and Its Athletic Program

In this section Michigan State University and its athletic programs are described, giving the setting the scope of the program for which the KBARS was constructed. It includes particulars about the university, the athletic department,

women's athletics, the head coaches, and the results of the ten sport seasons for 1977-1978.

Michigan State University, East Lansing, Michigan, is a land grant university founded in 1855. In 1977-1978 its student enrollment was 43,749. Its intercollegiate athletic department is a non-academic unit and largely self-supporting. Its athletic director, who reports directly to the president, is in charge of the men's athletic program as well as the program--men's and women's--as a whole. The current athletic director is in his second year at the university and holds a doctorate in education administration. Three assistant athletic directors report directly to him. Each one is responsible for one of the following areas as it affects athletics: Academics, Business Affairs, and Women's Athletics.

The men's athletic program is a member of the Big Ten conference and the National Collegiate Athletic Association. This program consists of 14 sports. Basketball, football and ice hockey are classified as revenue-producing sports, while baseball, cross country, fencing, golf, gymnastics, lacrosse, soccer, swimming/diving, tennis, track and field, and wrestling are the "non-revenue" sports. Athletic grants are awarded in all sports except in fencing, lacrosse, and soccer.

The women's athletic department is headed by the Assistant Athletic Director--Women's Athletics. She is responsible for virtually every aspect of the women's program which began in 1972. An administrative assistant, one secretary, ten head coaches, and

twelve assistant coaches comprise her staff. The university is a member of the Association of Intercollegiate Athletics for Women and its corresponding Regional and State associations.

The program is funded by sources from the athletic department and other funds from the university. An athletic grant program began in 1976 and is being phased in year by year. It is one of the smallest grant programs of universities in the Big Ten structure. There is a large disparity between the men's and women's programs especially in the areas of budget and facilities. Efforts are underway to bring the program into compliance with Title IX.

Ten teams and twelve seasons comprise the women's athletic program: field hockey, fall golf, volleyball, cross country, basketball, swimming/diving, gymnastics, indoor track and field, tennis, spring golf, outdoor track and field, and softball. The majority of the athletes are from Michigan. All teams have extensive schedules which take them to many parts of the United States to face top-ranked teams. Since 1972 women's teams have won numerous Big Ten, State, and Regional titles in addition to placing frequently in the top ten in National competition. In 1976 the softball team was the national champion.

The goal of the program is to help each female athlete to develop as much of her potential as possible. Consequently, all the sport seasons are considered to be equally important. This factor is reflected in the allocation of monies for both budget and athletic grants for each team.



Basketball. The head coach of the basketball team was in her second year at the university. Prior to this time she taught and coached for five years at a Michigan high school. She has competed in basketball and softball at the national level. In addition to being a coach, she is a specialist in the Department of Health, Physical Education and Recreation at Michigan State University.

The basketball team was the defending State and Regional champion. It successfully defended its State title but lost the Regional crown during the 1977-1978 season. A large part of its season schedule was played outside of Michigan in the Midwest and Northeast. Its final season record was 22-7.

Cross Country. The cross country coach was head coach for that sport and assistant coach for the track and field team. He is a doctoral student in sociology at Michigan State University. As an athlete he competed at the national level in the long jump and in cross country. He has coached the cross country team for three years. In 1976 the squad placed third in National competition. In the current season the team placed first in four invitational meets, third in Regional competition, and seventeenth in National competition.

Field Hockey. The field hockey coach was in his first year as head coach of the sport after having been the assistant coach for two years. He came from Thailand having competed and coached at the international level for that country. He has coached Thailand's

national men's team as well as a women's university team. The latter won three consecutive national titles. Currently he is a doctoral student in Student Personnel Services at Michigan State University.

The 1977-1978 hockey season was highlighted by a win over a team ranked fifth nationally. The team compiled a 9-3-3 record against teams from Michigan and Big Ten institutions. In State competition the team did not place.

Golf. The golf coach was in her sixth year as head coach of the team. Her husband coached the men's team at the university. She also worked as a professional at a local golf course. As an amateur she played golf at the national level defeating the late Mildred (Babe) Zaharias in match play in 1946.

Including the current season, the golf team captured six consecutive regional titles and five straight Big Ten championships. Each string of titles represents the complete history of the respective tournament. In addition to winning these titles, the squad won five invitational tournaments and placed eleventh nationally in 1977-1978.

Gymnastics. The head coach of the gymnastics team was in his first year in this position after being the assistant coach for two years. Prior to coming to Michigan State University he was assistant gymnastics coach at the University of Massachusetts. As a competitor he has won regional titles in tumbling and on the trampoline. In addition to his coaching responsibilities he is an assistant

professor in the School of Hotel, Restaurant and Institutional Management holding a doctorate in business administration.

The 1977-1978 gymnastics team not only competed in the Midwest, but also faced five of the top ten teams in the country. The team itself was ranked ninth nationally. Its dual record was 7-3. It won the State meet and placed first in the Big Ten, second in Regional competition, and twelfth at the National championship.

Softball. The softball team is the only team in the women's athletic program which has won a national title (1974-1975). During the 1977-1978 season the team embarked on a spring trip to New Mexico and a goodwill tour to Belize. Its season record was 22-35. It placed first in the Big Ten and second in the State tournament. The team entered the Regional tournament but did not place.

The coach of this team was in her third year at Michigan State University and was also an instructor in the Department of Health, Physical Education and Recreation. She had coached one year prior to coming to this institution. She is enrolled in a doctoral program in biomechanics at the University of Connecticut.

Swimming and Diving. The swimming coach was in her first year as coach of the Michigan State swim team and also in her first year as head coach at any institution. She was assisted by a diving coach and an assistant coach. She was a former national competitor and has served as assistant coach at Wayne State University. She

held a dual appointment with the Department of Health, Physical Education and Recreation as a specialist. The team had a 5-4 dual meet record and placed seventh in the Big Ten championship. There was no State or Regional championship event in this sport. One diver qualified for the AIAW nationals.

Tennis. The tennis team was coached by a coach who was in his first year at Michigan State University. He was a tennis teaching professional at a tennis club. He has held various positions related to tennis for a total of seven years. The team consisting primarily of underclassmen had a dual record of 4-8. It placed second in the State tournament. It entered the Regional event but did not place.

Track and Field. The track and field program consists of a winter indoor season and a spring outdoor season. The head track and field coach was in her second year in this position. In addition to coaching, she is also the administrative assistant to the women's athletic director. She still competes in marathons and set a world record in this event in 1971. She has a master's degree in physical education and has coached at a small college in Minnesota as well as at Oklahoma State University.

The track and field team placed second at both the indoor and outdoor Big Ten track and field meets and although six athletes qualified for the National championship, the team did not place. During the indoor and outdoor seasons the team won five meets and

placed third in a meet with nationally ranked University of Tennessee and Pennsylvania State University. One of its athletes competed in the World University Games in Bulgaria.

Volleyball. The head coach in this sport was in her fourth year in this position. She is also an assistant professor in the Department of Health, Physical Education and Recreation. Prior to coming to Michigan State University she taught and coached at a small college in Michigan. During her four years at the university the volleyball teams have won the State and Big Ten championships twice, the Regional title once, and have placed in the top ten twice in National competition.

The 1977-1978 volleyball team consisted mainly of freshmen who posted a 7-34-2 record against competition in the Midwest and Southwest. It placed third in the State tournament.

## CHAPTER II

### REVIEW OF RELATED LITERATURE AND RESEARCH

The research and literature to be reviewed will be representative of two major areas. The first area will encompass a discussion of the traditional approach to the appraisal of coaching performance and of the qualities purportedly related to successful coaching behavior. This area is further broken down into four topics: (1) personality profiles of coaches; (2) athletes' perceptions of coaches; (3) beliefs concerning essential coaching qualifications; and, (4) current methods of evaluating coaching performance. The second major area to be reviewed will consist of a summary of the literature and research which applies to the development and use of behaviorally anchored rating scales.

#### Traditional Approaches to the Evaluation of Coaches

In this section the traditional approaches to the evaluation of coaches as well as the qualities related to successful coaching performance will be discussed. The following four topics will comprise this section: (1) personality profiles of coaches; (2) athletes' perceptions of coaches; (3) beliefs concerning essential coaching qualifications; and, (4) current methods of evaluating coaching performance.

Personality profiles of coaches. There have been several attempts to classify and/or study coaching behavior by means of personality assessments. In such assessments coaches' traits are compared with those of their athletes and/or general population norms. The majority of such studies involve male subjects at the college level.

Those investigators who compared college coaches with the general population have found several differences. Andrud (1970) administered the Guilford Zimmerman Temperament Survey to nineteen coaches attending a football clinic. He found that these college coaches possessed a significantly stronger drive to succeed and higher energy than the general population. College football, basketball, track and baseball coaches were the subjects in a study by Ogilvie and Tutko (1966). The investigators administered the Jackson Personality Inventory. The results indicated that the 64 subjects scored significantly higher than college males on the following characteristics: orderliness, organization, warmth, trust, dominance and emotional maturity. The subjects revealed a strong need to be on top, tended to blame themselves when things went wrong and had finely developed consciences. The authors concluded that the most pronounced personality trait of these coaches was their high amount of psychological endurance. The results of Albaugh's (1972) study of 24 college basketball coaches and 175 varsity athletes indicated that there was no significant

difference between the two groups in resentment even though their mean score was one standard deviation higher than that of 900 teachers. Sage (1974b) found that the 343 college coaches representing basketball, football, and track programs across the United States did not differ significantly from college students in Machiavellianism.

Both Hendry (1969) and Loy (1969) compared personality traits of swimming coaches with each other and with that of their swimmers. Loy concluded that the trait of venturesomeness distinguished innovative swimming coaches from traditional coaches in the same sport. Hendry administered the Cattell 16 Personality Factor Inventory to 56 swimming coaches and their swimmers. He found no significant differences between the athletes and their coaches nor between successful and unsuccessful coaches on any factor. He concluded that the following description was characteristic of most successful coaches:

an outgoing, stable individual who would to some degree dominate the sports situation and the athletes in his charge . . . highly intelligent, realistic, practical, confident and secure . . . inventive, willing to break with tradition and to make his own decisions (1969:303).

Hendry decided that this personality pattern typified both athletes and coaches making it a "sport specific personality" pattern.

Relatively few personality profile studies are based on data from female coaches. Clay (1974) found that 34 female college coaches were significantly more intelligent and suspicious than their 185 athletes while the latter were significantly more



assertive than their coaches. There was no difference between the profiles of the coaches and the norms for 30-year-old women. Neal and Tutko (1975) reported that female coaches whom they had studied were more emotionally mature, independent and toughminded than their noncoaching counterparts. Two of the coaches in the aforementioned Loy (1969) study were female. Loy stated that these two subjects did not possess any personality traits which differentiated them from the male coaches.

The data generated by these personality investigations were summarized by Ogilvie and Tutko (1966) and also Sage (1975). Ogilvie and Tutko concluded that, in general, coaches are high in traits which are necessary for achievement and low in those traits related to sensitivity and nurturance. Sage characterized coaches as individuals who possess a high need to achieve, a great amount of energy and an ability to reach set goals.

The applicability of the above results to assessment of coaching behavior is, however, limited. The instruments which were used in the cited studies were not designed for the sporting environment and do not take into account the goals and values inherent in athletic programs. The sampling procedures varied widely ranging from all coaches enrolled in a class (Ogilvie and Tutko, 1966) to those who returned questionnaires (Sage, 1975). In addition, not one trait has been shown to be essential to coaching success or unique to members of the coaching profession. Consequently it becomes necessary to use other means to describe coaching performance.

Athletes' perceptions of coaches. Another aspect of assessing coaches is by means of the perceptions of their players of the coaching performance. Danielson, Zelhart and Drake (1975) administered a modification of the Leader Behavior Description Questionnaire to 160 junior and senior high school hockey players to obtain descriptions of the coaching behaviors of their coaches. The investigators selected 57 of the most commonly cited behaviors for factor analysis and multidimensional scaling. The factor analysis procedure yielded twenty factors which accounted for only two-thirds of the variance. Consequently, multidimensional scaling techniques were used and resulted in eight dimensions of coaching behaviors. These dimensions were labelled by the researchers as follows: (1) competitive training, (2) initiation, (3) interpersonal team operation, (4) sociability outside of team functions, (5) public relations, (6) organized communication, (7) recognition, and (8) general excitement (disorganization). Encouraging team members to work as a team was the most frequently mentioned coaching behavior. On the basis of the scaling results and lists of behaviors, the investigators concluded that the most prevalent perception of coaches by their players pertained to communication of information.

Meyer (1972) compared the self-perceptions of twelve male head college basketball coaches with those of assistant coaches and players at large universities. The rating instrument was a semantic differential based on the following six factors: general evaluative, potency, oriented activity, stability, receptivity, and aggression.

Players rated coaches and assistant coaches while assistant coaches rated themselves as well as their head coaches. The latter rated only themselves. Meyer found that the players perceived the head coaches to be more stable than the assistant coaches. Assistant coaches perceived the head coaches more favorably on each factor than did the players or the head coaches themselves. Since the staff members in seven of these basketball programs did not complete the instrument, a comparison was made of players' perceptions of coaches who had and those who had not participated in the study. The ratings of head coaches involved in the study was significantly higher than those of nonparticipants on the oriented activity factor. Meyer concluded that: 1) coaching staffs were apprehensive about research involving coaches, 2) assistant coaches perceived head coaches more favorably than players, and, 3) head coaches were more stable than assistant coaches.

Percival (1971) used an open-ended questionnaire and a numerical ten-point scale to collect opinions from athletes regarding their attitudes toward coaches and from coaches with respect to their own images. His survey included 318 male and 64 female athletes from six levels of competition ranging from age-group to professional sport. Seventy per cent of the athletes who participated in the study had been rated as being outstanding on their respective teams. The coaching sample included 52 males and 14 females representing all levels of competition. Percival categorized the responses into four areas: personality, techniques and

methods, knowledge, and mechanics. All comments were classified as being positive, neutral or negative.

The average overall rating of coaches by the athletes was five with the range extending from zero to nine. The mean overall self-rating of coaches was eight while the range of their ratings extended from five to ten. In addition to this discrepancy, Percival found that the higher the level of competition, the more apt a coach and athlete tended to rate the coach higher. Yet at the professional level the average self-rating of coaches was nine while the mean rating of these coaches by their athletes was six. The comments made by the athletes support these differences since 66% were negative, 10% were neutral and 24% were positive.

The area which received the lowest number of positive comments (32%) was the personality area. The coaches, however, made the greatest number of positive comments about themselves (72%) in this area. In terms of general knowledge, athletes' comments were positive 39% of the time as compared to 59% by the coaches. Forty-four percent and 62% of the comments by athletes and coaches respectively were positive concerning the techniques and methods area. The athletes and coaches agreed most in the mechanics area as was evidenced by 58% positive comments by athletes and 68% by coaches. Percival also classified the sports into team and individual sports. Seventy-nine percent of the comments of the individual sport athletes were negative as compared to 53% by team sport athletes.

Percival summarized all the results. He indicated that a large number of athletes dislike their coaches and are able to articulate this dislike to others but not to the coach. There was less negative response by female athletes to female coaches than to male coaches. The former were found to be less dictatorial and more inspirational. Percival also concluded that individual sport coaches were less competent than their counterpart in team sports. The content of the comments indicated to Percival that certain mannerisms of coaches "turned off" athletes. The athletes wanted a coach who was tough and fair, a leader, and concerned with their problems. Percival concluded that the ultimate limit to success in coaching is determined by the coach's personality since it generates a great deal of negative response from the athletes. Percival, however, cautioned against a generalization of his results since his sampling procedures and instrumentation were "not scientific."

Gaintner (1976) also investigated the self-perceptions of coaches as compared with athletes' perceptions of coaches. His results were different from those of Percival. He developed and used a 40-item five-point numerical Likert type scale based on a review of the related literature and a pilot study. He subsequently administered it to coaches and athletes who were members of water-polo and swimming teams. The results showed that there was no significant difference between the two groups on any of the items. Three items which were rated highest overall were: (1) knowledge of

skills required for success in the sport, (2) ability to plan and conduct worthwhile and meaningful workouts, and (3) ability of the coach to win the athletes' confidence in him as a coach. The three items rated lowest overall were the following: (38) amount of previous coaching experiences in the sport, (39) amount of previous competitive experience in the sport, and (40) creation of a friend relationship between the coach and athlete. Gaintner concluded that both coaches and athletes perceived the coach similarly.

LaGrand (1970) used a ten-point semantic differential scale to investigate the range of responses of athletes to behavioral characteristics of their coaches and to compare the profiles of individual and team sport coaches. The sports of basketball, soccer, wrestling and tennis were each represented by 76 athletes from colleges and universities in Vermont. On the basis of a review of literature and a pilot study, LaGrand identified 14 coaching characteristics for assessment. The subjects were asked to rate their "best" coach on each concept. The 14 concepts and their resulting order of importance as assigned by the athletes were as follows:

1. knowledge of the sport
2. enthusiasm
3. willingness to give individual help
4. demands for hard work
5. ability to organize
6. methods of teaching
7. interest in each player
8. ability for personal demonstration
9. ability to inspire
10. understanding of you as an individual
11. use of discipline
12. personal appearance
13. use of humor
14. interest in your out-of-school activities

Subjects were also requested to give an example of an incident which illustrated each concept. This procedure yielded 1298 statements out of a possible 4256 (30.4% yield). These were all positive statements. According to LaGrand these statements stressed that coaches should be sensitive to players' needs, possess a thorough knowledge of the technical aspects of the sport, and give individual attention to team members regardless of their skill.

When LaGrand compared the traits of team sport coaches with those of individual sport coaches, he found no significant differences on eight of the concepts. The largest differences in scores between these two groups of coaches were found in the concepts "demands hard work" and "ability to organize." Team sport coaches scored higher in both these characteristics.

LaGrand made several conclusions based on the results. For this sample a definite hierarchy existed in players' perceptions of coaches. He postulated that the reason the tennis coaches were rated lowest of all four groups may have been due to the fact that tennis does not have a high status as a varsity sport in Vermont because of seasonal influences. Thus the individual sport results were confounded by this variable.

In a similar study, Clark (1975) used the semantic differential to assess the characteristics of female intercollegiate coaches whose teams participated in tournaments sponsored by the Midwest Association for Intercollegiate Athletics for Women in 1972-1973. A total of 137 basketball players, 143 volleyball

players, 65 gymnasts and 74 swimmers participated in her study. They rated their coaches on 12 characteristics. Clark selected these concepts on the basis of the results of LaGrand's study and a review of related literature. The characteristics and their resulting order of importance as assigned by the 419 athletes were as follows:

1. knowledge of the sport
2. ability to teach
3. knowledge of coaching techniques
4. sense of humor
5. talent for organization
6. ability to inspire
7. ability to communicate
8. ability to motivate
9. fairness in dealing with each player
10. personal appearance
11. interest in my out-of-school activities
12. understanding of me as an individual

Overall, all groups rated their coaches favorably. The swimming coaches were rated the highest of all four groups while basketball coaches were rated higher than volleyball coaches. Those in individual sports were rated higher than their counterparts in team sports on all but the following characteristics: "knowledge of the sport," "ability to teach," and, "talent for organization."

Clark concluded that the four groups of coaches did not possess common characteristics. The concepts which were rated lowest overall were those which involved the player and coach in a one-to-one relationship. The results paralleled those of LaGrand (1970) in that "knowledge of the sport" was rated highest by both samples and "personal appearance," "interest in my out-of-school activities," and, "understanding of me as an individual" were ranked



lowest. Neither study, however, established the importance of the lowest ranked concepts nor made comparisons with the self-perception of coaches.

Jones (1975) conducted a study to determine the importance of coaching behaviors as perceived by volleyball players. Her subjects were 133 participants representing 13 institutions in the Illinois State Volleyball Tournament in 1974. She used a critical incident questionnaire to elicit statements describing coaching behaviors. The resultant 1324 statements were categorized into 353 effective and 331 ineffective behaviors. Jones was able to establish a rank order of important coaching characteristics based on the number of incidents which were illustrative of a specific type of behavior:

1. has ability to analyze and correct skill errors
2. encourages players and team
3. treats all players equally
4. gives individual instruction
5. develops team and player self confidence
6. praises good play
7. gives individual compliments
8. conducts team meetings to work out problems
9. uses challenging drills
10. jokes and laughs with players
11. plays substitutes when team is well ahead
12. has confidence in the team
13. uses drills to develop skills
14. substitutes if a player is doing poorly
15. motivates the team members so they will perform well
16. stresses teamwork at all times
17. has a positive attitude toward the players and the team at all times

Jones concluded that these 17 concepts were critical requirements for volleyball coaches at the collegiate level.

Beliefs concerning essential coaching attributes. The previous sections have described various studies in which an attempt was made to assess the qualities of coaches by means of an instrument. In addition to the results of these studies, there is a plethora of opinion as to the basic qualifications of a coach. Most of the writers of the literature in the coaching area cite knowledge as an important coaching attribute. This includes an up-to-date knowledge of skills, strategies and techniques related to the specific sport as well as a thorough background in kinesiology, physiology, psychology and conditioning methods (Counsilman, 1968; Deatherage and Reid, 1977; Klafs and Lyon, 1977; Miller, 1974; Neal, 1969; Poindexter and Mushier, 1973; Sabock, 1973).

Another essential qualification of a coach is that she must be a teacher and educator assuming all the responsibilities of being a faculty member (Cratty, 1973; Frost, 1971; Ley, 1973; McKinney, 1970; Miller, 1974; Rice, 1959; Shirley, 1966). This also means that the coach must demand respect and create positive public relations in all her dealings with students, peers, administrators and the community as would any other staff member (Lawther, 1977; Sabock, 1973; Singer, 1972). Loyalty to the institution is an aspect of this facet also (Gallon, 1974; Purdy, 1973; Sabock, 1973).

A coach must possess administrative skills. She must adhere to stated school and conference policies and procedures as well as being able to manage a budget (Deatherage and Reid, 1977). She

must not only be able to organize practice but also to coordinate the use of facilities, equipment, and, schedules (Deatherage and Reid, 1977; Ralston, White and Wilson, 1975).

The coach must possess the "right character." According to Sabock (1973) this quality is the key to coaching success and its major facet is sensitiveness to people. Poindexter and Mushier (1973) include stability and patience as being essential character traits. Other qualities frequently mentioned are leadership, emotional control and the ability to make decisions (Deatherage and Reid, 1977; Edwards, 1973; Gallon, 1974; Lawther, 1977; Singer, 1972). Such a coaching personality should reflect a "sound" philosophy as evidenced by the coach's actions and ethical practices (Klafs and Lyon, 1977; Neal, 1969; Poindexter and Mushier, 1973; Sabock, 1973).

The coach should be able to put the concern of others, such as the welfare of her athletes, ahead of her own needs and concerns (Butts, 1976; Purdy, 1973). She must be able to communicate this concern, as well as her knowledge, to the players (Butts, 1976; Cratty, 1973; Deatherage and Reid, 1977; Klafs and Lyon, 1977; Neal and Tutko, 1975). By means of this communication she must be able to inspire her personnel (Ralston, White and Wilson, 1975; Resick and Erickson, 1975). In essence, a coach by means of communicative processes and utilization of psychological principles, should be a "guiding light" (Counsilman, 1968).

The aforementioned characteristics are those most frequently mentioned in conjunction with coaching attributes in the literature. Such characteristics are abstract concepts in many cases and thus are difficult to assess. In addition, their importance to coaching success has not been documented. Thus they are opinions rather than facts. An instrument to appraise these qualities would be more accurate if it were based on established qualifications rather than opinions.

Current methods for evaluating coaching performance. Few instruments for appraising coaches have appeared in the professional literature. Emery (1962) constructed a scale for this purpose and published the steps entailed in its development. He constructed items based on a survey of the coaching literature. These items were submitted to a five-man jury who rated each item on a 5-point scale. Any item which received a total of less than 15 points was eliminated. This procedure resulted in a total of 114 useable items. These were rated on a 5-point scale by a 20-man jury consisting of administrators, physical educators and coaches who had coached for at least 15 years. The three characteristics which were rated highest for their contribution to coaching success were: 1) teaching ability, 2) knowledge of the sport, and, 3) ability to motivate. Emery suggested that his scale be used by both coaches and administrators.

The method of construction of other existing scales as well as an analysis of the resultant data are not available. Their

authors contend that the main purpose of these scales is self-evaluation. Friedrich (1953) published a scale consisting of 45 items. A coach is rated on a continuum from 0 - 10 for each item. Each continuum is anchored by a phrase on the left, in the middle, and on the right. Twenty-one of the items deal with such personal qualifications as cheerfulness, unselfishness, and voice. The other items deal with coaching techniques and skills such as "ability to teach fundamentals," "ability to instill determination," and, "willingness to accept suggestions." Gallon (1974), Palmieri (1974) and Pflug (1974) all are authors of scales which consist of categories judged to be essential to coaching success which are assigned numerical values by the rater.

Even though the literature does not abound with rating scales there have been many suggestions about what should constitute coaching success. Many such opinions are centered around the won/lost record. Both Sage (1975) and Massengale (1974) stated that in general a coach is judged to be successful if s/he wins more games than s/he loses. Rabolvsky (1974) suggested that the best coaches are those with break-even records since such records eliminate hysterical extremes by the community and administration. Maclean and Wilner (1977) indicated that the Russians employ a unique system based on the won/lost record: if a team fails to win, the players are replaced, not the coach!

Others stated that coaches should be judged by different criteria than the won/lost record. Ley (1973) believed that such

a performance appraisal should be goal centered. She asserted that since teachers are judged to be successful based on the extent to which students achieve established objectives, then coaches should be evaluated on the degree to which players attain individual and group goals. Deatherage and Reid (1977) believed a coach should be evaluated according to the standards and philosophy of the university and athletic department. Such an evaluation should include both the students and administrators. Tharp and Gallimore (1976) quoted Wooden as saying that a coach must be assessed in terms of the personal and psychological development of each individual athlete rather than the team's overall record. By such a standard, Wooden judged Vince Lombardi a failure.

Gabe Paul (1977), current president of the New York Yankees, summarized most of the current standards for appraising coaches as he cited his seven criteria for hiring or firing a coach:

1. What was the team's won-lost record?
2. Does the manager work hard enough?
3. Is he emotionally equipped to lead the men under him?
4. Is he organized?
5. Is he prepared?
6. Does he understand human nature?
7. Is he honorable?

Obviously the coaching profession is in dire need of methods to evaluate coaches which are concrete, reliable, and valid!

Summary. The foregoing discussion illustrates the fact that there is agreement on some of the qualities essential to coaching success. Players and administrators all give high ratings to knowledge of the sport, ability to teach, and, the effective use

of coaching techniques and communication abilities. However, there is much disagreement as to the importance of other attributes. In addition, existing evaluations of coaching performance are written in global attribute terms instead of specific behaviors. If these attributes themselves are not concretely defined then the results of any such appraisal will be an unknown quantity. Consequently, any form used to evaluate coaches must define clearly the qualities to be rated as well as establishing the importance of these qualities to coaching success. It is necessary therefore to examine the research related to the development of an instrument which is behavior specific. Such an instrument is the behaviorally anchored rating scale known as BARS.

#### Development and Use of Behaviorally Anchored Rating Scales (BARS)

Development of BARS. Smith and Kendall (1963) were the originators of the behaviorally anchored rating scale (BARS). Their intent was to develop a scale which would increase inter-rater agreement as well as face validity for the rater. The scales were constructed for use by head nurses to rate the performance of staff nurses in a variety of hospitals and different working conditions. The BARS consisted of nine dimensions with approximately eight items per dimensions. Item reliability coefficients ranged from .972 to .999.

The Smith-Kendall procedure, with a few variations is the one which is typically used to develop a BARS. This procedure

includes three major steps: (1) item and dimension development, (2) retranslation, and, (3) final scale construction. To develop the items and dimensions the rater population or samples thereof submit titles of dimensions and statements which illustrate effective and ineffective performances. The statements are sorted into homogeneous task clusters which are matched with the dimension titles and definitions. Every attempt is made to retain the original language.

As part of the retranslation technique, samples of the rater population allocate statements to dimensions and assign numerical values to each statement. Dimensions are retained if a minimum number of items are assigned to it. Behavioral statements are retained if there is a minimum percentage of agreement by the sample on the dimension of assignment and their standard deviation does not exceed an established minimum.

The final scale is constructed on the basis of the data generated by the first two steps. The average of the values assigned to each statement during the retranslation technique is used to place the statements on a continuum to serve as anchors for each dimension. The statements are rephrased in "expectation" terminology but otherwise kept in their original language.

Such steps with minor variations were followed by several investigators. DeCotiis (1974) developed a six-dimensional BARS for patrolmen. He found that his instrument was not sensitive to average midrange behaviors. His median reliability coefficients of



the dimensions ranged from .83 to .98. Landy, Farr, Saul and Freytag (1976) also focused on the rating of police officers by peers and supervisors. Over 2000 police officers were involved at various stages of the scale's development. A peer and a supervisor rated each officer. The intraclass reliability coefficients for two raters per dimension ranged from .47 to .73. Landy et al. found that the resulting scale was very useful for counselling and feedback and could be used by those not involved in its construction. Cascio and Valenzi (1977) also developed a BARS for police officers involving 71 raters and 299 ratees. They found that neither the rater's nor the ratee's educational background and/or job experience significantly affected the results.

The BARS has been used in a variety of other situations. Fogli, Hulin and Blood (1971) developed a BARS for grocery clerks. They reported an average item reliability coefficient of .97. Landy and Guion (1970) constructed a BARS to assess motivation of engineers while Motowidlo and Borman (1977) assessed the morale of army units. The results of both studies indicated that the BARS procedure was applicable to these situations even though the inter-rater reliability coefficients ranged from low to medium. Goodale and Burke (1975) investigated the extent to which a BARS could be used for more than one type of performance appraisal. They developed a BARS to evaluate all hospital personnel except for nurses, doctors, and administrators. All 60 eligible employees from 15 departments were involved in the construction of the instrument. Goodale and

Burke concluded that since 10 dimensions were retained with six to seven items per dimension that the resulting instrument could be used to evaluate employees in different jobs in the same hospital. Norton, Gustafson and Foster (1975) attempted to establish the degree of management potential of trainees. They found that a BARS instrument could be used and that female raters were significantly less lenient than male raters.

The performance of faculty members has also been appraised by students with the use of BARS. Harari and Zedeck (1974) used the input from 213 students to develop such a scale. Thirty-eight students contributed 310 behavioral statements while 175 were involved in the retranslation process. The procedure resulted in 9 dimensions and 78 useable behavioral statements. The investigators reported that the resultant scale was concise but insensitive to midrange behaviors. They also reported that students were competent and mature raters. Kafry, Zedeck and Jacobs (1976) used the above scale to evaluate the teaching effectiveness of introductory statistics instructors and found that the results were similar psychometrically to those of Harari and Zedeck's scale which had been constructed for psychology instructors. Zedeck, Kafry, and Jacobs (1976) compared the BARS results to those generated by a checklist and a graphic rating. The BARS results correlated highest (.86) with an overall rating of the instructor. The investigators suggested that the raters be asked to state additional incidents for each dimension in addition to rating the instructor

on each dimension. This would enable the investigators to keep the form current and would enhance the counselling use of the scale.

Jordan (1976) used the BARS procedure to develop an instrument to evaluate the performance of university department chairmen and deans. He found that the techniques most commonly used to establish the reliability of tests were inadequate for BARS. Consequently he established the reliability of his scale in three different ways. He randomly divided the subjects who were part of his retranslation phase into two groups. Mean scale values assigned by each group to each statement were correlated to assess scale reliability. These coefficients of item reliability ranged from .982 to .997. The instrument was administered to a group of faculty members and deans from which data the intraclass coefficients were calculated. They ranged from .40 to .86 for the faculty members and from .69 to .93 for the deans. Jordan also calculated the coefficient of agreement for each scale using a nonparametric technique (Lawliss and Lu, 1972). All the coefficients of agreement were significant at the .001 level indicating that agreement within each group was not due to chance. Jordan concluded that the developed BARS was adequate but that it should not be used indiscriminately at other institutions. He suggested that further analysis include factor analytic procedures.

The BARS in comparison with other scales. Borman and Vallon (1974) compared the results of a summated scale with those of the

BARS developed by Smith and Kendall (1963). The resultant ratings of nurses yielded higher inter-rater coefficients on the BARS as well as less leniency errors and halo effects. These results however were confounded by the fact that the appropriateness of the Smith-Kendall scale for this situation was not established. This is vital since the scale was at least ten years old.

Dickinson and Tice (1973) compared the results of a BARS scale for evaluating firefighters with actual ratings of job performance based on peer nominations and peer and supervisors' ratings. Using the multitrait multirater matrix techniques, the authors concluded that the BARS' dimensions had low discriminant validity and a moderate degree of convergent validity as well as a high degree of rater bias. They also concluded that "high agreement about what constitutes effective performance does not guarantee substantial agreement about actual performance" (p. 437).

The results of comparisons of BARS with other scales in a variety of situations are equivocal. Peters and McCormick (1966) developed a BARS to appraise worker activity and to compare the results with those of a numerical rating scale. The mean inter-rater reliability coefficients for the BARS ranged from .89 to .97 and were higher than those of the numerical scales. Millard (1975) compared the results of BARS and graphic ratings of state government workers. He found that the factor analysis of the BARS data was clearer and showed greater dimension independence than that of the graphic rating. Campbell, Dunnette, Arvey and Hellervik (1973)

developed a nine-dimensional BARS to evaluate department store managers of a national chain. The results were compared to those of a numerical scale. The investigators found that the BARS data showed less leniency and halo error than the numerical scale. They also indicated that the BARS had greater utility than the numerical scale because the developmental procedures forced the raters to define effective performance in behavior specific terms.

Groner (1975) compared the data generated by BARS to that of an adjective anchored scale and of a checklist. The focus of the scales was on the evaluation of videotapes of interviews by college recruiters. The BARS data showed greatest convergent and discriminant validity but was as susceptible to rater differences as the other scales. Groner also concluded that the usual techniques to determine reliability are inapplicable to evaluate the reliability of BARS since reliability theory assumes that true error variance in ratings is random which is not the case in BARS.

Freidman and Cornelius (1976) compared the results of graphic ratings by seniors of ROTC instructors with the results of BARS. Both scales were developed by the conference method. A group ( $n = 6$ ) which had not been involved in either of the conferences also used both scales to rate the instructors. On the basis of the results, the authors concluded that rater participation in instrument development significantly increased convergent validity and decreased the halo effect and that there was no difference between the results of the graphic rating scale and those of BARS.

The results of BARS have also been compared with other scales in the college setting. Bernardin, Alvares and Cranny (1976) compared BARS data to those of a rigorously developed numerical scale. The scales were used to rate college instructors. The authors found no significant differences in discriminant validity between the two scales. The numerical scale results showed less leniency error and higher inter-rater reliability than those of BARS. Bernardin, Alvares and Cranny concluded that the psychometric superiority of the numerical scale to that of the BARS was due to the fact that the numerical scale had been developed from the dimensions and items which constituted the BARS. In addition, the 556 students who rated the 27 instructors completed both rating scales which may have contaminated the results.

In a study using the ratings by twelve raters of three psychology instructors, Burnaska and Hollman (1974) found that the use of BARS resulted in excessive halo which was less than the halo generated by a numerically anchored scale and a traditional rating scale. Slater (1976) also developed a BARS for use in student evaluation of college instruction. His purpose was to determine if leniency in rating was a function of scale format. An adjective anchored scale was compared with the BARS. Fifteen instructors were thus evaluated. Half of each class received one format, the other half the other scale. There was no significant difference between the overall means of the two scales. Thus Slater concluded that BARS did not reduce rater leniency error. If,

however, the five dimensions of Slater's BARS scale were truly independent then the overall mean rating may not have been an accurate assessment of overall performance. In addition it was not possible to determine what the true average ratings should have been. Perhaps the fifteen instructors were outstanding.

Summary of studies involving BARS. Procedures for the development of BARS were summarized. Studies which detailed the construction and use of BARS were reviewed in terms of psychometric properties and comparisons with results of other scales. The results of the comparative studies are inconclusive concerning the relative psychometric advantages and disadvantages of the BARS. According to Schwab, Heneman and DeCotiis (1975):

almost all of the comparative research involved the evaluation of a BARS with an alternative instrument using the same dimensions as BARS . . . This means that we know little about the relative value of BARS when compared to typical rating procedures where performance dimensions as well as scaling formats tend to be chosen in an ad hoc fashion. (p. 223)

In addition, the studies which compared halo effects and leniency errors typically only involved two groups so that it was impossible to determine the true rating. Consequently, the psychometric advantages and disadvantages of the BARS are still relatively unknown and need to be explored.

## CHAPTER III

### PROCEDURES

According to Smith (1976) the greater the extent to which performance ratings reflect actual performance, the more accurate the ratings. Thus the ideal rating instrument is one which is psychometrically sound and can be effectively used by untrained raters. The procedures in the development of a BARS incorporate both these criteria by involving potential raters and by applying standards for inclusion of items in the actual form. The goal of the BARS procedures is to establish a set of independent, unambiguous and important performance dimensions, each of which is accompanied by a set of behavioral examples which help to define the performance factor. Such procedures are detailed in this chapter as they apply to the construction of a BARS for the performance appraisal of coaches at Michigan State University. The procedures were organized into the following seven phases: (1) Item and Dimension Development; (2) Retranslation Procedures; (3) Analysis of Retranslation Results; (4) Construction of the KBARS; (5) Administration of the KBARS; (6) Statistical Analysis of the KBARS; (7) Psychometric Analysis of the KBARS.



### Phase I: Item and Dimension Development

The first stage in the development of KBARS was to evaluate coaching behaviors involved by (1) explaining the project to the potential rater population and requesting their cooperation and involvement in it; (2) collecting statements from the rater population and from the coaching literature which were illustrative of effective and ineffective coaching behavior; (3) editing and preparing these statements for use in the retranslation phase; and, (4) selecting and categorizing descriptions of coaching behaviors to form dimensions.

Orientation of the Rater Population. Since the accuracy of the BARS evaluation instrument is dependent on rater participation, the procedures were explained to different rater groups. Four weeks prior to the start of the fall term, letters were sent to members of the women's athletic coaching staff, to returning female athletes and to first-year scholarship athletes. These letters introduced and detailed the study and explained the roles of the raters in the development of the KBARS. A copy of the letter is in Appendix A.

At the first meeting of the coaching staff for 1977-78, the investigator explained the details of the study stressing the fact that the extent of the accuracy of the appraisals using the KBARS would be highly dependent on the cooperation and involvement of the coaches and athletes.

Collection of Behavioral Statements. All potential raters were asked to recall and write specific behavioral statements which illustrated effective and ineffective coaching behaviors. The recall method was used instead of direct observations for several reasons. First, direct observations would have been collected at the end of each season and to a certain extent, recall would have been employed anyway. Secondly, it was assumed that all raters had previous experience in athletics and consequently, with coaches and their behaviors. If any of the elicited statements reflected atypical behaviors, then they would be filtered out by the retranslation phase of the development. Thirdly, Campion, Greener and Wernli (1973) studied the efficacy of the recall method as opposed to the direct observation method in the development of BARS and found that the use of the recall method did not significantly detract from the accuracy of the instrument.

Statements which were descriptive of coaching behaviors were collected at fall team meetings, from replies to the initial letters, and from the literature about athletics. The form used to collect these statements was similar to that used by Flanagan (1954) and Jones (1975) in the collection of critical incidents. Examples of both an effective and an ineffective coaching behavior were cited on the form. The raters were requested to submit at least three effective and three ineffective examples of coaching behaviors they had actually seen or experienced. It was stressed that the statements should be illustrative of behaviors and not be statements of

attributes. Behavioral examples were not restricted to the sport(s) in which the raters participated nor to coaches at Michigan State University. A copy of the form used to collect these statements is presented in Appendix A as well as examples of the type of incidents and statements collected.

Refinement of Statements. The behavioral incidents collected as described above were refined for further use. Statements of attributes, such as "The coach has a nice personality" were eliminated as well as statements which were redundant and vague. Although the original language of each statement was retained as much as was possible, the incidents were edited to reduce their length and to eliminate grammatical errors and references to specific people. These data refinements were completed by the investigator with the help of three persons: another coach, a graduate student, and the women's athletic director.

Dimension Selection and Refinement. On the basis of their content, the refined statements were sorted into tentative homogeneous categories. These categories were assigned titles and descriptions and thereafter were considered to be dimensions. Dimension information was refined and edited on the basis of input from the women's athletic director and the performance categories suggested by Clark (1975), Jones (1975), and LaGrand (1970).

## Phase II: Retranslation Procedures

This phase of the study consisted of the retranslation procedures modified from those developed by Smith and Kendall (1963). For this process, a sample was selected from the rater population. These individuals assigned numerical values to the refined behavioral statements and allocated them to dimensions.

Selection of Sample. The sample involved in the retranslation procedures was partly selected by random sampling techniques and then augmented by designated others so that it reflected the composition of the population. The random sample involved the athlete and trainer subgroups. Two trainers were randomly selected from those who had been members of the training staff at Michigan State University for at least one year. Three athletes were drawn randomly without replacement from each team's roster. Fall sport rosters consisted of actual active certified participants. Winter and spring rosters consisted of athletes who were on athletic scholarship, had attended the team's fall meeting, and/or had been a varsity or junior varsity athlete in a given sport for a minimum of one year.

For this sampling process, the rosters of the diving and swimming teams were combined as well as those of the cross country and track and field teams. This was done for several reasons. The entire diving team was new and consisted of only two members. Also, the diving coach, a member of both the men's and women's athletic

programs was unwilling to take part in this phase of the development of the KBARS. Since all the cross country team members were also listed on the track and field roster and the head cross country coach was also the assistant track and field coach, these two teams were also combined for the random sampling procedures.

The ensuing random sample was augmented by the women's athletic director, the head coaches, and assistant coaches. All the head coaches except for the diving coach participated in this phase. Since every assistant coach, except the one in track and field, was new and since no team managers, graduate interns and/or statisticians had yet been selected, all assistant coaches who had been hired by September 1, 1977, were included in the sample. The head track and field coach also held the position of administrative assistant. Since her job did not include responsibilities in the area of personnel, she was classified as a coach rather than as an administrator.

Retranslation by the Sample. The job of the members of this sample was to allocate the refined statements to the dimensions and to assign values to each statement on a seven-point numerical scale. The literature pertaining to the construction of BARS is not consistent in terms of recommendations for the optimum number of points per dimensional scale. The range for such scales has been from five to nine points. Bernardin, LaShells, Smith and Alvares (1976) found no significant differences in the results from BARS with different

formats and suggested that rigorous development of the scales was of greater importance than the actual format. Dickinson and Tice (1977) suggested that too many scale points would result in a scale with large gaps and decrease the advantages of the use of BARS. On the basis of these suggestions and the fact that Harari and Zedeck (1974) developed a BARS for college students to rate professors using a seven-point numerical scale, such a scale was chosen for this study.

The athletes were notified by the investigator of their selection for this phase of the study. A list of the participating athletes was given to each coach. The investigator, a graduate student and the coaches then reminded each athlete to attend the retranslation session. The athletes and other members of the sample who could not attend this session were given the booklets individually and were requested to complete the retranslation procedure within one week and return them to the investigator.

Each member of the sample was given a retranslation booklet. The first part of the booklet described the general procedures and their purpose. This was followed by a description of the dimensions and a request to read them carefully. It was suggested that the dimension sheet be removed from the booklet for easy reference. The next section of the booklet consisted of the behavioral statements. The subject was requested to allocate each statement to one of the dimensions using letters A through L and then to assign a numerical value ranging from one to seven to indicate the extent of the

effectiveness or ineffectiveness of the behavior. Each rater was instructed to work independently. A list of the items and dimensions is presented in Appendix B.

### Phase III: Analysis of Retranslation Results

The data generated by the retranslation procedure were analyzed to assess the extent to which the sample as a whole as well as its various subgroups agreed on dimension allocation and item value. The standards for analysis were adapted from and based on those used by others in the construction of BARS.

Retention of Dimensions. Dimensions were retained if at least four items were assigned to them. This criterion was used by others who used a seven-point numerical scale in a BARS (Bernardin, Alvares and Cranney, 1976; Fogli, Hulin and Blood, 1971; Slater, 1976; and Zedeck and Blood, 1974).

Retention of Items. An item was retained if it met two criteria sequentially. The first criterion required a minimum of 60% agreement by the raters on dimension assignment (Bernardin, LaShells, Smith and Alvares, 1976; Jordan, 1976). Bernardin et al. (1976) investigated the results of employing different levels of percentage of agreement on the resultant BARS and found that the 60% agreement level did not generate significantly different data in the final ratings than those based on 80% agreement levels. On the basis

of their results they recommended a 50% to 60% agreement level since this allowed for the inclusion of more items on the scale.

Assuming at least 60% of the sample agreed on the dimension to which an item was allocated, then the item was retained if the standard deviation of the rating was equal to or less than 1.75 (Bernardin, LaShells, Smith and Alvares, 1976; Slater, 1976). The selection of 1.75 as the criterion level for the variability was based on the standard deviation criteria used in studies which utilized a seven-point numerical scale, which involved college students as raters, and/or were developed for use in a multitrait, multirater situation (Dickinson and Tice, 1973, 1977; Harari and Zedeck, 1974; Lawler, 1967). The option of relaxing this standard was considered for some items near the middle of the scales since the variability of such statements would tend to be larger than that of items at the ends of the scales (Jordan, 1976; Landy and Guion, 1976). If items for the same dimension had similar mean values, the statement with the smaller standard deviation was retained since this indicated more agreement on the mean value.

Between Group Agreement. It was necessary to determine to what extent it was possible to construct a single set of rating scales which could be used by the various subgroups of the rater population. Since the majority of the retranslation sample consisted of athletes, their assignment of item values was compared with that of the rest of the sample. The extent of the between group agreement



was established for each item by means of the chi square test of significance. Items were retained if the values assigned to them by the athletes and by the rest of the sample did not differ significantly at the .05 level of significance (Jordan, 1976).

Since the sample had not been selected solely by randomization techniques, it was also necessary to determine the extent to which value assignment was due to chance. Such analysis was based on the random subdivision with stratification of the entire sample into an Original Group and a Cross Validation Group (Jordan, 1976; Schneider, 1976; Smith and Kendall, 1963). The sample was stratified according to the number of head coaches, assistant coaches, support personnel, athletes and administrators. Each stratum was proportionally represented in each group. Between group agreement was established for each item by the use of the chi square test of significance. Items were retained if the values assigned to them by the two groups did not differ significantly at the .05 level.

The degree of stability of each item was assessed in order to obtain an internal reliability coefficient for each dimension. The mean scale values assigned to items within each dimension by the Original Group were correlated with those of the Cross Validation Group by means of the Spearman Rank Order Correlation method.

#### Phase IV: Construction of the KBARS

Items and dimensions which survived the retranslation procedures were used to construct the actual KBARS. Each dimension was

placed on a separate page while items were reworded into expectation terminology and then placed at their scale value in their respective dimension. An appraisal booklet was compiled to facilitate administration of the instrument.

Final Selection of Items and Dimensions. Items surviving the retranslation procedures were sorted into their respective dimensions. The statements were anchored on a seven-point scale on the basis of the mean value (to the nearest .25) assigned to them by the sample. Thus an item with a mean value of 5.83 was placed on the scale at 5.75.

Each dimension was written on a separate page. A vertical continuous scale was used so that the items depicting ineffective coaching behaviors were placed at the top of the page while the items illustrating effective coaching behavior were at the bottom. According to Morrow (1976) such ordering of items on the continuum, as opposed to the reverse order, tends to reduce leniency errors.

Each item was worded in expectation terminology, that is, underneath each dimension description was the phrase, "This coach could be expected to . . ." This required the raters to infer performance and alleviated the problem of using anchors which might not be an exact description of a coach's current performance (Dickinson and Tice, 1977; Harari and Zedeck, 1974). According to Smith and Kendall (1963) the use of such terminology is vital to the success of the scale since it minimizes rating errors and deliberate faking by the raters.

Appraisal Booklet. The appraisal booklet used by the rater in the evaluation of the coaching performance of the head coaches consisted of three parts. In Part I all of the dimensions were listed on one page. The raters were required to rate the coach on a seven-point numerical scale. A value of one (1) represented extremely ineffective or very poor performance while a value of seven (7) represented extremely effective or very good performance. The middle values (3-4) indicated average performance. According to Lawler (1967) and Zedeck, Kafry and Jacobs (1976), these global ratings tend to reduce leniency and halo errors in subsequent ratings. Thus, such ratings should increase the accuracy of the ratings in Part II, the actual BARS of the appraisal.

Part II required the raters to appraise the coach on each of the 10 dimensions on a seven-point numerical scale. Each dimension was listed on a separate page. The seven-point numerical scale for each dimension incorporated the behavioral statements as anchors to help the rater define the scale values. The raters had to judge which scale value best described the coach's past and/or expected behavior. Only one number could be chosen per dimension. The selected number was marked on the answer sheet.

The third part (Part III) of the booklet requested the rater to describe critical incidents and/or comment on behaviors by the coach which substantiated the numerical ratings. Such incidents assisted the investigator in determining the extent of the validity of the use of the KBARS for coaches in the variety of sports and the

validity of profiles generated by the scores from Part II. A coach who received high ratings in Part II should also receive mainly positive comments in Part III.

Answer sheets were used in the appraisal booklet as recommended by Landy, Farr, Saal, and Freytag (1976). They suggested that the use of separate answer sheets could reduce paper and printing costs, facilitate scoring, and decrease the amount of file space needed to keep the results in administrative offices. A copy of the booklet and the answer sheet are found in Appendix C.

#### Phase V: Administration of the KBARS

Team members, support personnel, the women's athletic director, one member of the coaching staff and the coach her/himself responded to Parts I through III at the end of a particular season. The team trainer supervised the appraisal by the athletes at one of the team's practices. The coach was requested to be absent during this appraisal as is customary procedure for all faculty evaluations at Michigan State University. Each coach completed a self-evaluation and was rated by the women's athletic director, by all support personnel associated with the team and by a peer who was selected by the coach.

Originally it was intended that every coach evaluate all the other head coaches but several factors limited the peer evaluation to one selected peer appraisal. First of all, four head coaches were in their first year of coaching at the university and found it

impossible to accurately rate the fall sport coaches. Secondly, coaches whose sport was in season experienced difficulty in observing each other in coaching action since there was a great deal of overlap in schedules for practices and events. Thirdly, the growth of the women's sports program at Michigan State University has dispersed practices and events for the different sports from one building to three. Thus some coaches rarely saw each other except at staff meetings. Consequently, each coach selected a peer from the head coaching staff who had had the opportunity to observe the coach in action at practices and events to rate her/him.

The completed KBARS ratings were scored by the investigator. A mean score and its standard deviation were calculated for each dimension for each coach. These statistics were reported for the athletes and staff as well as for the rater group overall. The comments and/or incidents from Part III of the instrument were presented along with each dimension and its statistics. The resulting profiles were given to the women's athletic director who discussed them with the respective coaches.

#### Phase VI: Statistical Analysis

The data were examined so that the common elements underlying coaching performance at Michigan State University could be examined. A correlation matrix was constructed to indicate the relationships between the dimensions and to reveal the psychometric properties of the KBARS. This was followed by a principal component

factor analysis. The Statistical Package for the Social Sciences (SPSS) was employed on the CDC 6500 computer. The eigenvalue was never less than one since all dimensions were considered to be equally important. Any extracted factors, therefore, would account for at least 10% of the variance (Rummel, 1970). In addition, although the goal of the construction of the KBARS was to create independent dimensions, it was acknowledged that such dimensions automatically have a basic or common variance since they were all related to the coaching performance of one person (Latham and Wexley, 1977). Consequently, the magnitude of the loading of each dimension on a factor had to be greater than .50 to be considered significant.

#### Phase VII: Psychometric Analysis of the KBARS

The usefulness of an appraisal instrument depends to a large degree on the extent of its psychometric properties (Dunnette, 1976; Schneider, 1976). Consequently, the degree of validity and reliability of the KBARS had to be established (subproblem 2). The studies involving the development of BARS have investigated the psychometric properties, if at all, in various ways with most of the emphasis on the establishment of reliability.

Response Set Bias. The results of the KBARS were examined for response set biases. A judgment of the tendency towards leniency was based on the size of the resultant mean ratings of Part II of the appraisal. The lower the mean, the smaller the leniency error

(Bernardin, Alvares, and Cranny, 1976). Halo effects in the ratings were determined by an examination of the correlations between dimensions. The higher these correlations, the greater the halo effect (Smith, 1976). The nature of the comments/incidents generated by Part III of the appraisal were used to judge the extent to which the leniency errors and halo effects were due to inadequacies of the instrument. Since the purpose of Part I of the appraisal was solely to reduce error response sets, the data generated in those procedures were eliminated from the psychometric analysis.

Reliability. Jordan (1976) stated that the parallel forms method and the split half method to determine reliability were inapplicable to BARS. Nunnally (1967, pp. 214-215) recommended that the test retest method to determine reliability not be used when only a small number of qualities were rated or when there were less than six months between the test and retest. Groner (1975) evaluated theories of reliability and found them to be inapplicable to most rating data because the theories assume that true variance and error variance are independent and that the latter is random. Both assumptions are not applicable to rating data. Because of such contentions, scale developers have examined the internal reliability of scales instead of the external reliability. Most investigators who have reported the reliability of BARS have done so only in terms of scale reliability. According to Schwab, Heneman and DeCotiis (1975) inter-rater reliability is as important, if not more so, as

scale reliability. Consequently, both methods of determining reliability were used in the present investigation.

The internal reliability of the KBARS was established in two ways. Scale reliability was based on the correlation between mean scale values assigned to the behavioral anchors of each dimension by the Original Group and the Cross Validation Group in Phase III of the KBARS. The resultant correlation coefficients reflected the degree of agreement on the values of the behavioral statements for each dimension. Inter-rater reliability was determined by the calculation of the intraclass coefficient as was suggested by Ebel (1951) and Safrit (1976). The resultant coefficients of correlation indicated the degree of agreement between the raters on dimension ratings.

Validity. Whereas there were no available rating scales with which to compare the results of the appraisal, other methods of validation were used. This was done in other studies using BARS. The instrument had, by the very nature of its development, a high degree of content validity (Dunnette, 1976; Smith, 1976). Kavanagh, MacKinney and Wolins (1971) and Lawler (1967) recommended that validity also be determined by an evaluation of convergent and discriminant validity. Dickinson and Tice (1977) Latham and Wexler (1977) and Smith (1976) suggested that the factor analytic approach be used rather than the multitrait multirater matrix as recommended by



Campbell and Fiske (1959) since the latter is affected by small fluctuations in the correlation coefficients. Consequently, the results from a factor analysis and analysis of variance of the data were used to determine the extent of the convergent and discriminant validities in this study. Rater agreement on dimensions indicated the extent of convergent validity. The degree of dimension independency, that is, low intercorrelations and low loadings between dimensions indicated the degree of discriminant validity (Zedeck and Blood, 1974).

## CHAPTER IV

### FINDINGS AND DISCUSSION

In this chapter the data generated by the different phases of the KBARS are presented. The results are compared to those of other studies in which BARS were developed. The extent of the psychometric properties of the KBARS are discussed as well as its feasibility in measuring common elements of coaching performance in the women's athletics program at Michigan State University.

#### Phase I: Item and Dimension Development

This phase of the construction of the KBARS to assess the performance of coaches involved in the women's athletic program at Michigan State University 1977-1978 involved the collection of behavioral statements and the construction of dimension descriptions. The potential raters were oriented to the study and were asked to write critical incidents/statements. These statements and incidents were then prepared for use in the retranslation phase. Dimension descriptions were constructed by the investigator on the basis of the submitted statements and incidents as well as the literature dealing with coaching.

Collection of Behavioral Statements. All persons who would eventually be using the KBARS to evaluate coaching performance were contacted and asked to submit behavioral incidents and statements which exemplified effective or ineffective coaching behaviors. Ninety-two athletes, nine coaches, two assistant coaches, one trainer and the women's athletic director submitted statements and/or incidents. The latter totalled 623 of which 58% were labelled "effective." Other developers of BARS instruments have also found that the rater population tended to submit slightly more effective than ineffective incidents (Jordan, 1976; Harari & Zedeck, 1974; Slater, 1976).

The 623 items were sorted into tentative homogeneous categories. For each of these categories a one-sentence description was written. These categories were refined and relabeled as dimensions. A total of twelve dimensions were selected for use in the construction of the KBARS. The descriptions of the dimensions, as well as their code names, were as follows:

- A. Behaviors concerned with scheduling and organizing home and away contents (Scheduling)
- B. Behaviors concerned with the organization and content of practices (Practice)
- C. Behaviors concerned with team selection and coaching philosophy (Selection)
- D. Behaviors which illustrated personality facets of the coach (Personality)
- E. Behaviors concerned with the setting of team rules, discipline and team control (Discipline)

- F. Behaviors concerned with skill development in practice and the use of these skills in contests (Skill)
- G. Behaviors concerned with the performance of the athletes in actual competitive experiences (Meets)
- H. Behaviors which affected the athlete psychologically and provided motivation and support for her (Support)
- I. Behaviors related to communication with the team and promotion of team work (Communication)
- J. Behaviors which denoted professional characteristics (Professionalism)
- K. Behaviors which denoted a knowledge of the sport and coaching techniques (Knowledge)
- L. Behaviors concerned with public relations and recruiting (P. R./Recruiting)

Each of these dimensions had at least 50 behavioral statements tentatively assigned to it with the exception of Dimension L--Public Relations and Recruiting. Consequently, the women's athletic director and the investigator created 20 additional statements which seemed to fit into this category. Jordan (1976) indicated that each dimension should have at least 30 items assigned to it in this phase.

The investigator, the women's athletic director, a coach, and a graduate student then examined the statements in each category and eliminated those which were vague, replicated, redundant and/or failed to illustrate specific behaviors. Incidents were rewritten into useable statements. The resultant statements were edited to reduce their length, to eliminate grammatical errors and to obliterate references to specific people. Many statements were duplicates especially those assigned to Dimension D--Personality, Dimension

H--Psychological Support, and Dimension I--Team Communication. Of the original 623 items, 299 statements survived this procedure (48%).

In the literature dealing with the development of BARS, there is little mention of the percentage of items which usually survive the refinement process and/or the minimum number which are necessary for use in the retranslation procedure. Harari and Zedeck (1974) developed a BARS for use by college students. They originally collected 310 incidents of which 78 (25.5%) statements and nine dimensions were used in the final instrument. Slater (1976) in a similar type of study used 53 items and five dimensions in the retranslation process. His final instrument consisted of five dimensions and 34 (64%) items. Jordan (1976) constructed a BARS to evaluate department chairpersons. His retranslation procedure was based on nine dimensions and 237 statements. All nine dimensions and 72 (30%) items were used in the final instrument. On the basis of these results, it would seem that the final instrument should possess about eight items per dimension. It may also be assumed that approximately two-thirds of the items could be lost during the retranslation procedure. If all the dimensions in the current investigation were retained, the final instrument would contain approximately 96 items requiring an initial item pool of at least 288 statements. Consequently, the total of 299 statements was judged adequate. These statements are presented in Appendix B.

## Phase II: Retranslation Procedure

Selection of Sample. The sample selected for the retranslation process consisted of athletes, head and assistant coaches, trainers and the women's athletic director. Three athletes were randomly chosen from each team's current roster. Not all of the selected athletes were able to participate which resulted in a total of 25 athletes representing the nine sports being involved in this phase. The breakdown by sport is given in Table 1. Since 44% of the nine sports in the program are team sports and the percentage of team sport athletes in the sample was 44%, the sample was judged to be representative of all female athletes at Michigan State University 1977-1978.

The sample of athletes was augmented by two randomly chosen trainers, nine head coaches, six assistant coaches, and the women's athletic director. Thus, the total sample consisted of 43 members.

Retranslation by the Sample. The retranslation phase took 1 1/4 to 4 hours; the average time was two hours. Jordan (1976) commented on the fact that the length of time involved in completing the retranslation procedure was a drawback in the construction of his BARS instrument. Many of those involved in the current investigation felt that the procedure was too time consuming.

## Phase III: Analysis of Retranslation Results

The data generated by the retranslation procedure were analyzed to assess the extent to which the sample as a whole, as well

Table 1  
Retranslation Sample

Sport	Athletes	Head Coach	Assistant Coach	Other	Totals (n = 43)	
					Athletes	Staff
<u>Individual/Dual</u>						
Golf	2	1	0	0	14 (56%)	7
Gymnastics	3	1	0	0		
Swimming/Diving	3	1	1	0		
Tennis	3	1	0	0		
Track and Field	3	1	1	0		
<u>Team</u>						
Basketball	3	1	1	0	11 (44%)	8
Field Hockey	3	1	1	0		
Softball	2	1	1	0		
Volleyball	3	1	1	0		
<u>Other</u>						
Trainers	-	-	-	2		3
Athletic Director	-	-	-	1		
Totals (n = 43)	25	9	6	3	25	18

as its subgroups, agreed on item allocation and values. The standards for analysis were based on and adapted from those used by others in the construction of BARS.

Retention of Dimensions. A dimension was to be retained initially if at least four items were allocated to it. All 12 dimensions met this criterion. The number of items assigned to each dimension are presented in Table 2. If an item was assigned to two or more dimensions by an equal percentage of the sample, then it was given a proportional value so that the total number of items still summed to 299 as Table 2 indicates. For example, statement 30: "The coach's directions are short and to the point" was assigned modally by 29% of the sample to Dimension J--Professionalism and by another 29% of the sample to Dimension K--Knowledge. Thus in terms of the number of items assigned to these dimensions, each dimension gained half (.5) an item.

The modal dimensions were Dimension D--Personality and Dimension H--Support with 45.5 and 56 items assigned respectively as Table 2 reveals. The dimensions with the least number of statements assigned to it was Dimension L--P. R./Recruiting. This repeated the phenomenon which had been observed in Phase I; that is, that Dimensions D--Personality and H--Support contained the greatest number of critical incidents whereas Dimension L--P. R./Recruiting had the least. Perhaps Dimension L overlapped with other dimensions. The popularity of Dimensions D and H illustrates the importance of



Table 2  
Number of Items Surviving Each Criterion

Dimension	Number <sup>a</sup> Assigned	Agreement ≥ 60%	s > 1.75	Between Group		Total			
				Agreement <sup>b</sup>		Eliminated		Retained	
				1 <sup>c</sup>	2 <sup>d</sup>	Number	Percent	Number	Percent
A. Scheduling	8.0	6	6	6	6	2.0	25.00	6	75.00
B. Practice	16.0	12	12	12	12	4.0	25.00	12	75.00
C. Selection	36.0	12	12	12	12	24.0	66.67	12	33.33
D. Personality	45.5	12	12	12	12	33.5	73.63	12	26.37
E. Discipline	20.5	12	12	12	12	8.5	41.46	12	58.54
F. Skill Development	13.0	1	--	--	--	13.0	100.00	0	0.00
G. Meets	17.0	0	--	--	--	17.0	100.00	0	0.00
H. Support	56.0	26	25	24	24	32.0	57.14	24	42.86
I. Communication	23.5	8	8	8	8	15.5	65.96	8	34.04
J. Professionalism	26.5	11	11	10	10	16.5	62.27	10	37.73
K. Knowledge	31.0	25	24	23	23	8.0	25.81	23	74.19
L. P. R./Recruiting	6.0	4	4	4	4	2.0	33.33	4	66.67
Totals	299.0	129	126	123	123	176.0	58.86	123	41.14

<sup>a</sup>Includes those assigned modally to more than one dimension

<sup>b</sup>Failure to reject H<sub>0</sub> at .05

<sup>c</sup>Athletes vs. Rest of Sample

<sup>d</sup>Original vs. Cross Validation Group

the personality of the coach and the mental support s/he gives to individual athletes.

Retention of Items. To be retained a statement had to be assigned to the same dimensions by at least 60% of the sample. Of the 299 items, 129 (43.14%) met this criterion as is shown in Table 2. Neither Dimension F--Skill Development nor Dimension G--Performance in Competition had the minimal number of items (4) surviving this criterion. Thus, these two dimensions and their corresponding items were eliminated. The aspects of performance illustrated by these dimensions must have overlapped greatly with those of the other dimensions.

Each of the items which met the dimension allocation criterion in the remaining ten dimensions was retained if its standard deviation was less than or equal to 1.75. Two items did not meet this criterion reducing the item pool to 126 statements (42.14%) as Table 2 illustrates.

The summary data of the items which survived these two criteria are presented in Table 3. The percentage of the modal frequencies per dimension ranged from 60.46% to 100%. The greatest agreement on the assignment of items was in Dimension A--Scheduling in which the median modal frequency was 95%. Thus, even though only six items in this dimension survived the retranslation process, the sample showed high agreement in placing these items in this dimension. The median frequencies also revealed that if the criterion for percentage allocation would have been 70% as in several

Table 3

## Dimension Statistics for Retained Items (n=123)

Dimensions	Percent Assignment		Standard Deviation		Total		Scale
	Range	Median	Range	Median	Retained	KBARS*	r
A. Scheduling	76.74 - 100.00	95.00	.75 - 1.61	.97	6	5	.83
B. Practice	60.46 - 97.67	76.74	.66 - 1.63	1.03	12	11	.95
C. Selection	60.46 - 93.02	67.44	.58 - 1.67	1.40	12	10	.98
D. Personality	60.46 - 88.37	66.00	.67 - 1.38	1.13	12	12	.91
E. Discipline	60.46 - 93.02	80.23	.84 - 1.68	1.33	12	12	.90
H. Support	60.46 - 88.37	66.28	.58 - 1.47	1.01	24	16	.92
I. Communication	60.46 - 88.37	67.44	.65 - 1.67	1.12	8	8	.89
J. Professionalism	60.46 - 69.77	66.28	.53 - 1.72	.96	10	9	.88
K. Knowledge	60.46 - 97.67	76.74	.59 - 1.68	1.01	23	17	.97
L. P. R./Recruiting	60.46 - 76.74	73.26	.96 - 1.30	1.07	4	4	1.00

\*This column depicts the actual number of items used in the KBARS (n=104).

other BARS procedures (Schwab, Heneman & DeCotiis, 1975), many more items would have been eliminated.

The majority of the items survived the standard deviation criterion as Table 2 indicates. The standard deviation of the remaining items ranged from .53 to 1.72. Dimension A--Scheduling and Dimension J--Professionalism had the lowest median standard deviation. This indicated that the sample agreed most on the values assigned to items in these two dimensions as compared to those in the other eight. Not only did the sample show a high degree of agreement on which items belonged to Dimension A--Scheduling, but also on the value of each allocated item. It is possible that the scheduling and organizing of home and away contests are essentially administrative functions which could be defined concretely. The reverse could be true of Dimension D--Personality and I--Communication which lost the greatest percentage of items. The sample disagreed as to which items belonged under these dimensions. Perhaps these dimensions represented aspects of coaching performance which were difficult to illustrate behaviorally.

Between Group Agreement. Judgments obtained from the sample were examined to determine the extent of agreement between certain subgroups of the sample by means of the chi square test of significance (Siegel, 1954). Yates' correction was used wherever the expected cell frequency was less than five. If the expected cell frequency was less than one, the cell was combined with the adjacent

cell. These techniques were followed to ensure that no more than 20% of the cells had expected frequencies less than five and that the test was as powerful as possible. The null hypothesis tested was that the values assigned each item by the various subgroups did not differ significantly. Items were eliminated if there was a significant difference at the .05 level in value assignment.

When the assignment of values by the athletes was compared with that of the rest of the sample, differences were significant for three items. Thus, on the whole, there was not much difference in values assigned to items by the athletes than by the staff. The resulting KBARS could, therefore, be used validly by athletes and staff.

Between group agreement was also examined for the Original Group and the Cross Validation Group. Differences in value assignment were significant for two items. This indicated that on the whole members of the sample agreed on value assignment and thus disagreement was due to chance.

There were two items which showed a significant difference at the .05 level for both comparisons. Thus, a total of three items were eliminated from the item pool as a result of the criteria pertaining to between group agreement. Since the reduction in the item pool was small as a result of these procedures indicating that the various sample members agreed on the values of the items, then the resultant KBARS could be used by the types of raters represented by the sample; that is, the subgroups of the rater population would not

need different instruments to assess the performance of the ten head coaches in the women's athletic program at Michigan State University 1977-1978.

Dimensions and Items Surviving All Criteria. The number of retained dimensions was judged to be sufficient. The loss of the two dimensions was interpreted as indicating that there was an overlap among the 12 dimensions. The infrequent assignment of items to Dimension L--Public Relations/Recruiting may signify that many coaching behaviors related to the other dimensions are also those which are vital to recruiting and public relations. In other words, a coach's relationship with her/his athletes may be a part of public relations and recruiting. Since the dimension still possessed the minimum number of items to be retained, it was kept as part of the instrument. The number of items which survived all criteria for retention is presented in Table 3. One hundred and twenty-three (41.14%) of the 299 retranslated items survived all the criteria used to screen the behavioral statements. Summarized on the basis of the order in which the criteria were applied, the item pool reductions occurred as follows: 129 (43.14%) remained after the criterion regarding percentage assignment to dimensions; 126 (42.14%) items remained after the criterion for standard deviation was applied; 123 (41.14%) of the items survived after the values assigned to statements by different subgroups were compared.

Comparison of the number of items surviving the retranslation process in the present study with those surviving the same phase in other BARS is difficult. Reasons for such difficulty are that (1) the retranslation information is not always presented, (2) the initial pool of dimensions differs from BARS to BARS, and (3) the resultant BARS were used for a variety of occupations. A few studies involving the construction of BARS for use by college students to rate faculty members do provide information for comparison with the present study. Harari and Zedeck (1974) used 54 students in a retranslation procedure involving nine dimensions and 310 statements. All nine dimensions survived the process along with 33.77% of the items which resulted in an average of approximately eight anchors per dimension. The majority of the items which were eliminated were those which failed to meet the dimension allocation criterion.

Keaveny and McGann (1975) found that all 13 dimensions and 74 (27%) of the statements survived the process. Slater (1976) retained all five dimensions and 64.15% of the original 53 items. Jordan (1976) also developed a BARS for the college setting but utilized faculty members for the retranslation process. Of the original items, 51% were retained, as well as all ten dimensions. In all of these studies, the final average number of anchors per dimension ranged from 6 to 12. In the current investigation, 10 dimensions and 123 items survived the process so that approximately 12 items were available per dimension. Such a number was judged to be adequate for the construction of KBARS.

The greatest reduction in the item pool came as a result of the dimension allocation criterion. Dimension D--Personality lost the greatest percent of items as Table 2 showed. This indicated that even though a large number of items was originally assigned to this dimension, the agreement on which specific items belonged in this dimension was low. A similar phenomenon was also reported by Harari and Zedeck (1974) and Jordan (1976). In the current study, this phenomenon, in addition to the fact that only a few items were lost as a result of other criteria, indicated that the retranslation sample had greater difficulty categorizing behaviors than evaluating them. This could have been due to dimension ambiguity and/or overlap. It is also possible that the aspects of coaching performance which were examined were highly related.

Scale Reliability. The mean scale values assigned to items by the Original Group were correlated with the mean values assigned to items by the Cross Validation Group to assess the degree of stability in mean values assigned to the anchors used in the construction of the KBARS (Jordan, 1976; Smith and Kendall, 1963). Values in Table 3 are Spearman Rank Order correlation coefficients computed on the mean values for the statements in each dimension. The coefficients were large (Nunnally, 1967) and ranged from .83 to 1.00. These data indicated that the retranslation sample was consistent in its assessment regarding the level of performance represented by the behavioral statements used to anchor each dimension.



Few investigators have reported scale reliabilities for BARS. Fogli, Hulin and Blood (1971) constructed a BARS to appraise performance of checkers at grocery stores. They reported scale reliabilities ranging from .97 to .99 for nine dimensions. Smith and Kendall (1963) reported similar coefficients in the construction of a BARS for rating nursing performance. A BARS for rating department chairpersons had scale reliabilities ranging from .982 to .997 (Jordan, 1976). Burnaska and Hollman (1974) reported reliability coefficients ranging from .81 to .97 for a 13 dimension BARS for rating college faculty. For a similar type of instrument, Bernardin, LaShells, Smith, and Alvares (1976) found the reliability coefficients ranged from .76 to .92. The scale reliability coefficients reported in the current study fell within the range reported by other investigators especially by those in which a BARS was used in the college setting.

The reliability coefficients discussed above should not be used to determine the overall reliability of the KBARS since they tend to be somewhat spurious. Their value could be raised if items were chosen which have small standard deviations and whose means are relatively far apart. An increase in the number of subjects involved in the retranslation procedure, as well as a decrease in the number of items per dimension, would also tend to raise the correlation coefficients. All of these factors, with the exception of the choice of anchors with small standard deviations, are not related to scale effectiveness. Consequently, the psychometric

analysis of the KBARS included a calculation of the intraclass coefficient for each dimension. These coefficients are reported in Phase VII.

#### Phase IV: Construction of the KBARS

On the basis of the results of Phase III, the judgment was made that a BARS could be constructed which could be used by all the members of the rater population to rate the coaches in 10 different sports. The items and 10 dimensions which survived the retranslation process were used to construct the KBARS. Each dimension was placed on a separate page. The statements were anchored on a seven-point scale on the basis of the mean value (to the nearest .25) assigned to them by the retranslation sample. Not all the items surviving the retranslation process were used in the KBARS as Table 3 indicated. If items had similar means the statement with the smallest standard deviation was chosen. Nineteen items were not used so that the KBARS consisted of a total of 104 statements, thus averaging about 10 statements per dimension. This selection process did not affect the scale reliabilities. The resulting KBARS is presented in Appendix C. As seen, the problem of the mid-range gap existed. For most of the dimensions, there were no items available with mean values ranging from 3.0 to 4.0 which met all but the standard deviation criterion. Thus, no statements were added to fill the mid-range gaps.

The entire appraisal instrument consisted of three parts. In Part I the raters were required to rate the coach on each of the 10 listed dimensions. A value of one (1) represented extremely ineffective or very poor performance, while a value of seven (7) represented extremely effective or very good performance. The middle values (3 to 4) indicated average performance.

Part II required the raters to appraise the coach on each of the 10 dimensions by means of a numerical value ranging from one (1) to seven (7). The seven-point rating scale for each dimension incorporated the behavioral statements as anchors to help the rater define the scale values. The raters had to judge which scale value best described the coach's behavior and/or expected behavior. Only one number could be chosen per dimension. The selected number was marked on the answer sheet. Part III of the KBARS requested the rater to comment and/or give examples of behaviors which substantiated the ratings assigned in Part II.

#### Phase V: Administration of the KBARS

Ratings by the Population. The KBARS was given to team members, support personnel, the women's athletic director, one member of the head coaching staff, and to the coach her/himself at the end of a sport season. The number of raters varied from sport to sport as Table 4 indicates. A total of 173 ratings was compiled of which 128 were completed by athletes and 45 by staff members. The ratings by the athletes represented 81.53% of those competing on

Table 4  
Size of Population Using KBARS

Sport	Staff	Percent Return	Athletes	Percent Return
Basketball	6	100.00	12	100.00
Cross Country	4	100.00	14	100.00
Field Hockey	4	80.00	16	84.21
Golf	3	100.00	8	100.00
Gymnastics	3	75.00	10	100.00
Softball	5	83.33	10	56.56
Swimming/Diving	5	100.00	12	46.15
Tennis	4	100.00	5	56.56
Track/Field	6	100.00	30	100.00
Volleyball	5	100.00	11	100.00
Total	45	92.50%	128	81.53%

varsity teams during the 1977-1978 season. Four of the ten teams had less than 100% returns. Of the ratings returned by the staff, 10 were completed by the women's athletic director and 10 were self-ratings. Twenty-five support personnel used the KBARS to evaluate head coaches. Of those staff members involved with specific teams and thus eligible to rate a coach, 92.50% completed the ratings. Three assistant coaches failed to complete the ratings on their respective head coaches.

The completed KBARS ratings were scored by the investigator. For each coach, the average score and its standard deviation were

calculated for each dimension. These statistics were reported for the athletes and staff as well as for the group overall. The comments and/or incidents from Part III of the instrument were presented along with the dimension information to the women's athletic director who used the resulting profiles for counselling and feedback to the ten head coaches. Since such profiles were confidential information, they are not presented in this study.

Discussion. The number of completed ratings was sufficient to proceed with the evaluation of coaching performance using KBARS. According to Rummel (1970) the ratio of the number of cases to variables should be as large as possible so that there is no restriction on the maximum number of orthogonal factors that can be extracted. In this study, there were 10 dimensions and 173 ratings resulting in a seventeen to one ratio of cases to variables. It was concluded, therefore, that the factors which were extracted were nonrandom and based on stable differences among correlations (Aleamoni, 1976).

The major impediment to achieving a 100% return rate was the coaches' attitudes toward the evaluation. Some were very resistant to allocating part of practice time to evaluation and/or ensuring that the support personnel completed the ratings. This was the first year that an evaluation system was employed which involved ratings by different groups. In previous years only the women's athletic administrator rated each coach. There have been few occasions in

athletics where coaches were rated on performance by athletes and other groups. Many coaches have, therefore, not been socialized into accepting ratings nor are they willing to provide time for athletes to complete them. As a multirater multitrait rating system, such as the current KBARS, becomes an integral part of each coaching season, the resistance by the coaches should fade and a 100% return rate should be achieved.

#### Phase VI: Statistical Analysis

Correlation Matrix. The correlations among the dimensions ranged from .31 to .72 and were all positive. The matrix is presented in Table 5. The Personality dimension consistently correlated the highest with the other nine dimensions, while the Scheduling dimension and the P. R./Recruiting dimension correlated the lowest with the others. Thus, the personal characteristics of the coach were strongly related to the coach's philosophy, methods of providing support for each athlete, communication with the team, professionalism, and team discipline.

Factor Analysis. The correlation coefficients reported in Table 5 were subjected to a factor analysis procedure so that the grouping patterns of the dimensions could be observed. Dimensions which were highly related should group together into a factor. The principal component factor model was applied to the data. The resulting factor and its factor loadings are presented in Table 6.

Table 5  
Correlation Matrix of KBARS Dimensions

Dimensions	Sched.	Prac.	Se1.	Pers.	Disc.	Supp.	Comm.	Prof.	Know.	P. R./Rec.
Scheduling	--									
Practice	.41	--								
Selection	.38	.42	--							
Personality	.47	.42	.63	--						
Discipline	.48	.51	.58	.56	--					
Support	.41	.32	.58	.72	.48	--				
Communication	.44	.39	.49	.66	.57	.62	--			
Professionalism	.44	.38	.53	.66	.52	.65	.56	--		
Knowledge	.40	.61	.41	.49	.41	.39	.33	.45	--	
P. R./Recruiting	.52	.39	.41	.58	.48	.46	.52	.49	.31	--

Table 6  
Factor Analysis of KBARS Dimensions\*

Dimension	Factor Loading	h <sup>2</sup>
Scheduling	.66	.44
Practice	.65	.42
Selection	.77	.59
Personality	.86	.74
Discipline	.69	.48
Support	.76	.58
Communication	.77	.59
Professionalism	.72	.53
Knowledge	.64	.41
P. R./Recruiting	.57	.32
Eigenvalue	5.10	5.10
% Variance	100.00	100.00

\*Principal Component

The factor loadings are coefficients of correlation representing the relationship between each dimension and the factor. For example, the Scheduling dimension had a factor loading of .66 indicating that the relationship between this dimension and the factor was significant if the factor loading criterion of .50 is used. The Personality dimension correlated highest with the factor since its factor loading was .86, whereas the P. R./Recruiting dimension had the lowest factor loading of .57. The median factor loading was .705.



In this solution, each communality ( $h^2$ ) is the square of its respective factor loading and represents the variance of the dimension which is shared by the other dimensions. For example, 44% of the variance of the Scheduling dimension is shared with the other nine dimensions. The sum of the communalities equals the total amount of variance which was accounted for by the factor which was 5.10 as Table 6 indicates.

The eigenvalue reveals the amount of variability which was retained by the factor and equals the sum of the squared factor loadings. Since the eigenvalue of 5.10 equalled the sum of the communalities, all or 100% of the variance was accounted for. Consequently, no further rotations of the data were needed. The factor was named the KBARS Factor.

Discussion. The KBARS factor analysis indicated that the instrument functioned as a unity. This increases the utility of the instrument. It permits a summation of the ten dimension scores so that a total can be obtained for each coach. This sum would be an index number which could be considered to be a rough estimation of a coach's performance during the season. Care should be taken, however, to use such a total score in conjunction with a profile of all ten dimensions since the latter did not load equally on the KBARS Factor.

The high degree of relationship shown between the Personality dimension and the KBARS Factor illustrated the importance of the

behaviors related to this dimension. This phenomenon was evident in Phase I and Phase III of the study also. In the former, the Personality dimension contained the greatest number of critical incidents while in the latter it was assigned the second highest number of items. This also illustrates the need for a BARS type of instrument. A large percentage of the items assigned to this dimension in Phase III did not survive all the criteria for inclusion in the KBARS. Those in the retranslation sample agreed as to which items belonged in this dimension, but disagreed as to the values assigned each item. They were able to categorize the behaviors, but had difficulty evaluating them on a numerical scale.

The results concerning the Personality dimension paralleled some of the conclusions found in the literature pertaining to coaches. Percival (1971) contended that the personality of the coach was the ultimate limiting factor to successful coaching performance. Other investigators (Hendry, 1969; Loy, 1969; Ogilvie & Tutko, 1966) attempted to determine the personality type of successful coaches. The present study revealed by means of the behavioral anchors, the type of personal characteristics which denoted effective and ineffective performance.

The low loading of the P. R./Recruiting dimension on the KBARS Factor also followed a pattern evidenced throughout the construction of KBARS. It was the only dimension which had to be augmented with additional incidents in Phase I and which had only the minimum number of incidents needed to survive the retranslation

phase. This may be explained by the fact that recruiting is a newly emerging dimension in the coaching of female athletes at the college level. AIAW regulations have been very restrictive in this regard so that the amount of observable recruiting has been negligible. Most recruiting in the women's athletic program at Michigan State University was done by correspondence and not visible to many of the raters. In rating their coaches on this dimension, many of the scholarship athletes in essence were rating themselves since they were recruited to come to Michigan State University. Perhaps they felt that if they rated a coach low on this dimension, it would be a reflection of their athletic ability. If a coach were rated low on this dimension, it might be interpreted that the athletes on the team were not symbolic of "good" recruiting.

#### Phase VII: Psychometric Analysis

The data were examined so that the psychometric properties of the KBARS could be established. The descriptive statistics from Phase VI were used to determine the extent of the response set bias as well as the degree of validity and reliability of the instrument.

Response Set Bias. The data of the KBARS dimensions were examined for response set bias. The means, standard deviations, ranges, and correlations which are presented in Tables 7 and 8 illustrate the extent to which the KBARS was affected by such response errors.

Table 7

## Summary Data of KBARS Dimensions by Rater Groups

	Self			Administrator			Support Personnel		
	$\bar{X}$	s	Range	$\bar{X}$	s	Range	$\bar{X}$	s	Range
Scheduling	6.20	0.79	5 - 7	5.20	0.92	4 - 6	5.36	1.55	2 - 7
Practice	6.40	0.52	6 - 7	5.10	0.99	3 - 6	5.60	1.32	2 - 7
Selection	5.60	0.97	4 - 7	5.20	1.23	3 - 7	5.32	1.60	2 - 7
Personality	5.80	1.03	4 - 7	5.30	1.34	3 - 7	5.84	1.49	1 - 7
Discipline	5.30	1.77	2 - 7	5.20	0.79	4 - 6	5.24	1.16	2 - 7
Support	6.20	1.23	3 - 7	5.30	1.49	2 - 7	5.68	1.52	2 - 7
Communication	5.80	1.32	3 - 7	4.80	1.14	3 - 6	5.36	1.47	2 - 7
Professionalism	6.40	0.70	5 - 7	4.90	1.45	3 - 7	5.72	1.54	1 - 7
Knowledge	6.30	0.67	5 - 7	6.00	0.94	4 - 7	5.88	1.36	2 - 7
P. R./Recruiting	6.00	0.82	5 - 7	4.90	1.10	3 - 6	6.04	0.98	4 - 7
	Athletes			Staff			All Raters		
	$\bar{X}$	s	Range	$\bar{X}$	s	Range	$\bar{X}$	s	Range
Scheduling	5.64	1.36	2 - 7	5.51	1.32	2 - 7	5.60	1.35	2 - 7
Practice	5.26	1.32	2 - 7	5.67	1.19	2 - 7	5.35	1.29	2 - 7
Selection	5.07	1.46	1 - 7	5.36	1.38	2 - 7	5.14	1.44	1 - 7
Personality	5.60	1.47	1 - 7	5.71	1.36	1 - 7	5.63	1.44	1 - 7
Discipline	5.37	1.62	1 - 7	5.24	1.23	2 - 7	5.33	1.52	1 - 7
Support	5.24	1.71	1 - 7	5.71	1.46	2 - 7	5.35	1.66	1 - 7
Communication	5.20	1.68	1 - 7	5.33	1.38	2 - 7	5.22	1.60	1 - 7
Professionalism	5.71	1.46	1 - 7	5.69	1.44	1 - 7	5.70	1.45	1 - 7
Knowledge	5.54	1.56	2 - 7	6.00	1.15	2 - 7	5.65	1.47	2 - 7
P. R./Recruiting	5.69	1.40	1 - 7	5.78	1.06	3 - 7	5.71	1.32	1 - 7

Table 8  
Medians and Ranges of Inter-Dimension  
Correlations of KBARS

Dimensions	Median Correlations	Range
Scheduling	.46	.41 - .49
Practice	.46	.38 - .61
Selection	.54	.41 - .64
Personality	.59	.46 - .74
Discipline	.50	.44 - .58
Support	.49	.39 - .74
Communication	.55	.37 - .66
Professionalism	.50	.40 - .66
Knowledge	.44	.30 - .61
P. R./Recruiting	.45	.30 - .59

According to Bernardin (1976) leniency errors can be detected in two ways. First of all, the lower the mean, the smaller the leniency error. The extent of the leniency errors can also be determined by the number of scale points actually used by the raters. The overall average ratings for the coaches per dimension were in the upper third of the scale ranging from a low of 5.14 for Selection to a high of 5.71 for P. R./Recruiting (Table 7). Since the values of 3 to 4 were labelled as "Average Performance" on each dimension, the coaches as a whole demonstrated above average performance on each dimension. These high values either indicate that the raters tended to be lenient and/or that the ten coaches actually did perform well in these dimensions.

As the ranges of the ratings indicate (Table 7) the coaches, when rating themselves, used the fewest number of scale points and primarily those values which were at the upper end of the scale. The lower half of the scale was used for only three dimensions (Discipline, Support, and Communication). The first scale point (1) was not used at all. This group also marked only the top two scale points in rating themselves in the Practice dimension.

All the other groups used all seven scale points. However, these raters, on the average, predominantly marked scale points ranging from three to seven.

The greater the rater's ability to distinguish among the dimensions of the KBARS and thus rate the ratee differentially on each dimension, the smaller the halo effect (Bernardin, 1976). In addition, the greater the variance of the ratings, the smaller the halo effect. The values of the standard deviations for the KBARS dimensions indicate that there was a great deal of variability among the ratings as a whole (Table 7). These parameters for the entire population ranged from 1.29 (Practice Dimension) to 1.66 (Support Dimension). Thus the KBARS did not tend to elicit the halo effect when the ratings were examined for the entire group of raters.

The self-ratings and those by the administrator did show a halo effect. In the former, six of the dimensions had a variance of less than one while in the latter there were four such dimensions.

This result could possibly have been a function of sample size since both the self-ratings and the administrator's ratings numbered ten.

The extent of the halo effect can also be assessed by the amount of dispersion among the interdimensional correlation coefficients (DeCotiis, 1977). Table 8 indicates the range of these coefficients, as well as the median correlation coefficient for each dimension. In general, there was quite a degree of dispersion among the coefficients. The median correlation coefficient for the entire instrument was .495. Others who have investigated the psychometric properties of BARS have reported median correlation coefficients for the entire instrument ranging from .32 (Keaveny & McGann, 1975) to .73 (DeCotiis, 1977).

The effect of the response set biases was difficult to assess since comparative data on coaching performance were not available. According to the definitions of leniency errors and halo effect, the KBARS indicated a high degree of leniency, especially for certain subgroups, and a moderate halo effect. However, it is possible that the ten coaches performed their job in an outstanding manner and truly merited the high ratings. The coaches, in their ratings of themselves, showed the greatest amount of leniency. This result paralleled those found in studies involving self-ratings by college faculty (Centra, 1972; Doyle, 1975). It is possible that once the coaches had access to the ratings of the other subgroups that their concept of their coaching performance might be altered and thus their self-ratings.

The halo effect was judged to be moderate. Since all of the dimensions were related or linked to one person (the ratee) in each rating set, a certain amount of correlation between the dimensions should have occurred.

Reliability. Since there is no ultimate method to determine the extent of the reliability of an instrument (Safrit, 1976), two methods were used to determine the extent of the reliability of the KBARS. The scale reliability coefficients ranged from .83 to 1.00 per dimension and were calculated as part of the retranslation procedure. They are reproduced in Table 9 and discussed in the section dealing with the retranslation procedures. They reflect the degree of agreement on the anchor values.

The intraclass coefficient of correlation was calculated for each dimension based on a two-way analysis of variance of the data from the ten dimensions by raters and ratees. According to Safrit (1976) and Ebel (1951) if it is known that the different groups of raters were using different frames of reference in their ratings, then between-rater variance should be removed from the error term before the intraclass coefficient is calculated. However, if the raters or judges should be using the same standard for judging performance, the between-rater variance component should be included. Although there was a significant difference among the groups of raters in this study, the between-rater variance was included in the calculation of the intraclass coefficient since the



Table 9  
Reliability Coefficients for KBARS

	Scale r	Inter-Rater R
Scheduling	.83	.79
Practice	.95	.72
Selection	.98	.73
Personality	.91	.87
Discipline	.90	.77
Support	.92	.80
Communication	.89	.61
Professionalism	.88	.83
Knowledge	.97	.75
P. R./Recruiting	1.00	.78

theoretical purpose of the KBARS was to give all raters the same criteria for judging coaching performance. The use of this error term produced a conservative estimate of between-rater agreement. Its use was also justified by the fact that this coefficient of correlation tends to be a biased estimate if there are less than 200 subjects involved in the analysis (Safrit, 1976). The resultant coefficients are presented in Table 9. The values range from .61 for the Communication dimension to .87 for the Personality dimension. The median intraclass coefficient was .775.

The scale reliability coefficients illustrated the extent of the agreement among those involved in the retranslation procedure

concerning the values of the anchors of each dimension. The KBARS was judged to have high scale reliability.

The intraclass or interrater reliability coefficients were lower than the scale reliability coefficients. Few investigators who have constructed BARS have reported interrater reliability coefficients. Jordan (1976) reported intraclass coefficients ranging from .40 to .93 for ratings by college faculty of deans and chairpersons. According to Nunnally (1967) the minimal level of the reliability of new measures is .60. Thus, the KBARS showed acceptable to high internal reliability.

Validity. The KBARS, by nature of its construction procedures, had a high degree of content validity (Dunnette, 1976; Smith, 1976).

According to Kerlinger (1973) content validation is a judgmental matter based on the extent of the sampling adequacy of the instrument (p. 458). Both the item development phase and the retranslation phase of the KBARS were attempts to ensure that the final dimensions and their respective scale anchors were adequate representative samples of coaching performance and behaviors. The 173 KBARS ratings of the coaches were an accurate assessment of coaching performance as the comments generated by Part III of the instrument indicated. Thus, the assumption that the KBARS had high content validity seemed justified. The extent of the convergent and discriminant validities of the KBARS were examined utilizing the

data from the correlation matrix (Table 5) and the results of a one-way analysis of variance (Table 10).

According to Zedeck and Blood (1974), intercorrelations between dimensions indicate the degree of discriminant validity of BARS. As Dickinson and Tice (1977) observed, it is almost impossible to obtain a high degree of discriminant validity of a BARS since the rating for each dimension represents part of the same whole, that is, the rater. As the factor analysis results revealed, the KBARS functioned as a unity thus affecting discriminant validity in an inverse manner. Since the interdimension correlation coefficients were generally moderate in size ranging from .31 to .72 (Table 5) with the median valued at .48, the degree of discriminant validity was judged to be low to moderate.

Rater agreement indicates the extent of convergent validity (Dickinson & Tice, 1977). Consequently, the results of a one-way analysis of variance were examined to see if the four different groups (self, administrator, athletes, and support personnel) rated in a similar manner on the KBARS as a whole. The results indicated that there was a significant difference among the factor scores of the four groups ( $p \leq .01$ ). The Least Significant Difference Test was then used to detect which groups differed significantly at the .05 level (Table 10). There was no significant difference among the ratings by the three staff groups. The staff as a whole, however, rated the head coaches significantly differently than did the athletes.

Table 10  
Analysis of Variance of Groups by Factor Scores

Source	df	SS	MS	F	p
Between Groups	3	10.72	3.57	4.26	.0063
Within Groups	169	141.79	.84		
Total	172	152.51			

Least Significant Difference Test (Factor Scores)			
Rater Groups	Administrator	Support Personnel	Athletes
<u>Self</u>			
Mean Difference	.1400	.2177	.7104*
Critical Value	.5768	.4863	.4388
<u>Administrator</u>			
Mean Difference	--	.0777	.5704*
Critical Value	--	.4368	.3831
<u>Support Personnel</u>			
Mean Difference		--	.4927*
Critical Value		--	.2827

\*Significant at the .05 level.

Consequently, the degree of convergent validity for the KBARS was high for the staff and low for the rater population as a whole.

The greater the extent of agreement by the different raters on the ratings, the greater the degree of convergent validity. According to Schwab, Heneman and DeCotiis (1975), one of the advantages of a BARS is that the anchored behaviors can give different raters the same reference points which enhances the degree of convergent validity. There were differences between the mean ratings by the coaches themselves, by the administrator, and by the support personnel, but these differences were not significant. This lack of significance indicated that the KBARS in this case fulfilled its purpose of providing the same reference points for those with different perspectives. The trainers and assistant coaches, as well as the head coaches themselves, were the only ones who were involved with the teams every single day of the season. The peer group and the administrator viewed the coaches' performance mainly at competitive events. They rarely saw the coach in practices and other situations. Yet, these different groups rated in a similar manner. Thus, KBARS was judged to have a high degree of convergent validity for the staff.

The significant difference between the ratings by the athletes and by the staff as a whole indicated a low degree of

convergent validity for the rater population as a whole. It also substantiated the trend found in student ratings of college instructors (Doyle, 1975). There were only two available studies in the coaching literature in which athletes and coaches perceptions of coaches were compared. Gaintner (1976) found no significant difference between the scores of athletes and coaches on a 40-item Likert-type inventory designed to rate coaches. Percival (1971) found that athletes and coaches perceived the coach differently in three areas: personality, techniques and coaching methods, and knowledge. Thus, the low convergent validity of KBARS may have reflected the manner in which college students and faculty perceive instructor performance. It may also have reflected an inadequacy of the instrument. Judgment must be reserved until athletes' and coaches' perceptions of coaching behaviors can be investigated further.

Since there was no criterion measure available to validate the KBARS, a judgment of its overall validity must be made to answer subproblem 2. Its high content validity indicated that it measured what it purported to measure, that is, coaching performance. The comments made by the raters in Part III substantiated this conclusion. These are discussed in the next section.

The divergence between the ratings of the staff and the athletes may be a desirable phenomenon since the involvement of these two groups with the coach differed and thus would tend to produce two different perspectives. On the whole, the staff

viewed only the product, that is, the whole team in action, while the athletes were involved in the day-to-day process of working towards this product. On the basis of the foregoing discussion, the KBARS was judged to have adequate validity to be used as an instrument to assess coaching performance of the 10 head coaches in the women's athletic department at Michigan State University.

Part III of the KBARS. The third part of the evaluation booklet requested the rater to comment and/or give examples of behaviors which justified the ratings for each dimension. These comments and/or incidents were primarily used for counselling and feedback by the women's athletic administrator.

The answers were classified as to whether they were "comments" or "incidents" and whether they were positive or negative. A tabulation of these results appears in Table 11. Not every rater responded to Part III and some raters reacted only to certain dimensions in this part. The data showed evidence of a trend. The Knowledge dimension evoked the most responses, while the P. R./ Recruiting dimension received the least. The majority of the responses were positive and in the "comment" category.

The replies to Part III indicated the need for an instrument such as the KBARS. The majority of the responses were general comments so that a coach and/or administrator would have had a difficult time accurately assessing performance and setting new goals based solely on these responses. The majority of the comments were

Table 11  
Classification of Comments/Incidents of Part III

	Comments		Incidents		Total
	+	-	+	-	
Scheduling	76	2	14	10	102
Practice	72	12	8	2	94
Selection	66	6	10	2	84
Personality	104	2	4	4	114
Discipline	68	2	4	10	84
Support	100	2	10	0	112
Communication	94	2	8	0	104
Professionalism	76	0	8	0	84
Knowledge	140	0	4	4	148
P. R./Recruiting	62	0	8	4	74
Total	858	28	78	36	1000

of the global attribute variety, e.g., "She communicates well with the team," "She knows her stuff," "Professionalism? I don't know if he ever played as a pro!" These types of comments again point out the necessity of an instrument such as the KBARS. If the behaviors anchoring the scales for each dimension are used as reference points, they can provide guidelines for coaching behavior. Positive global comments, however, deal mainly with the status quo; without a frame of reference they can be meaningless.

The Part III of the KBARS generated 114 examples of behaviors. The majority of these were positive. The administrator felt that these incidents were very helpful in evaluating the coaches in



conjunction with the KBARS results. They should be preserved and retranslated so that they can be used to update the KBARS and possibly help to eliminate some of the gaps in some of the dimensions.

The P. R./Recruiting dimension received the lowest number of comments/incidents which corroborated the pattern evidenced throughout the development of KBARS concerning this dimension. Comments such as the following were typical: "I don't know what she does in this regard." This echoes the assumption made earlier; that is, that in conjunction with AIAW rules, much of the work pertaining to this dimension is "invisible," since it is done by correspondence.

The fact that 93.40% of the comments and incidents were positive supported the contention concerning the average scores per dimension. It was stated that the magnitude of these scores could be attributed to leniency error and/or outstanding performance by the coaches. The prevalence of the positive comments/incidents support the latter conclusion. Thus, the leniency effect may not have been as large as was originally thought.

### Summary

In this chapter the results of the seven phases of the KBARS development and analysis were presented and discussed. An analysis of the results indicated that the KBARS is a viable multi-trait multitrait model for appraising coaching performance of the ten head coaches in the women's athletic program at Michigan State University. The use of the instrument generated ten dimensions

which were common elements of coaching performance. The KBARS possessed a high degree of content validity and scale reliability, moderate discriminant validity and interrater reliability, and a low degree of convergent validity. The coaches' profiles generated by the ratings paralleled those generated by the comments in Part III. This further substantiated the validity of the instrument and confirmed the fact that the leniency error and halo effect did not significantly alter the results. Thus, the results of the statistical and psychometric analysis of the KBARS answered the two sub-problems stated in Chapter I.

## CHAPTER V

### SUMMARY AND CONCLUSIONS

In this chapter the results of the administration of the KBARS are summarized. The feasibility of assessing coaching performance using the KBARS and the applicability of the KBARS results for coaching are discussed. Recommendations for use of the KBARS are presented, as well as suggestions for future study.

#### Summary

The purpose of this study was to construct a multitrait multirater model for the appraisal of coaches in the women's athletic program at Michigan State University in 1977-1978. The research was carried out in seven phases.

Phase I of the study consisted of item and dimension development for the Knoppers Behaviorally Anchored Rating Scale (KBARS). A total of 623 incidents depicting effective and ineffective coaching behaviors were obtained from 105 athletes, coaches, administrators, and trainers. These 105 people were all members of the population which completed the final ratings. The incidents were sorted into twelve dimensions. Items which were vague, replicated, redundant and/or failed to illustrate specific behaviors were eliminated. Of the statements, 48% (299) survived this procedure. Each dimension was next defined to make it as independent of the other dimensions as possible.

The retranslation procedure constituted the second phase of the study. A sample consisting of 43 raters-to-be was selected so that all the subgroups of the rater population were proportionally represented. The members of the sample allocated each statement to a dimension and assigned the statement a numerical value from one to seven with respect to its effectiveness or ineffectiveness.

Phase III consisted of an analysis of the retranslation results. The standards for analysis were based on and adapted from those used by others in the construction of BARS. Ten dimensions and 123 of the items (41.13%) met all the criteria for inclusion. The resulting ten dimensions were: (1) Scheduling; (2) Practice Content; (3) Team Selection; (4) Personality; (5) Team Discipline; (6) Psychological Support; (7) Communication; (8) Professionalism; (9) Knowledge; and (10) Public Relations/Recruiting. The items and dimensions were then incorporated into the actual KBARS, the construction of which was the fourth phase of the study.

The KBARS consisted of three parts. Part I required each rater to rate the coach on each dimension on a numerical scale without the use of behavioral anchors. The purpose of this rating was to reduce the halo effect. Part II required the raters to rate the coach on the ten dimensions using the behavioral anchors as a guide. Part III asked the raters to comment and/or give examples of behaviors by the coach for each dimension which substantiated the ratings given in Part II.

KBARS was administered to each team and relevant staff members at the end of its sport season. This was the fifth phase of the study. A total of 173 ratings were completed which included 128 by athletes and 45 by staff members. The ratings by the athletes represented 81.53% of those competing on varsity teams during the 1977-1978 season.

In the sixth phase of the research, the data generated by the 173 ratings were analyzed. Both the correlation matrix and factor analysis showed similar results. The intercorrelations among dimensions ranged from .31 to .72. The Personality dimension overlapped the most with the other dimensions, while the P. R./Recruiting dimension overlapped the least. The ten dimensions formed one factor accounting for 100% of the variance. Thus, the KBARS functioned as a unity.

The last phase of the KBARS development included a psychometric analysis of the instrument itself. A high degree of leniency, especially for the self and peer ratings, and a moderate halo effect were shown. The extent to which these were errors was unclear since the nature of the comments generated by Part III indicated that the coaches truly "deserved" the high ratings.

The extent of the reliability of the instrument was assessed in two ways. The scale reliability was calculated for each dimension and depicted the agreement among those in the retranslation sample on the values assigned to behavioral anchors. These coefficients ranged from .83 to 1.00. The intraclass coefficients of

correlation were calculated to indicate the extent of between-rater agreement on dimensions. These coefficients ranged from .61 to .87. On the basis of the magnitude of both sets of coefficients, the KBARS was judged to possess satisfactory internal reliability. External reliability was not assessed since current methods for obtaining such correlation coefficients are inappropriate for use on a rating scale such as a KBARS (Groner, 1975; Nunnally, 1967).

The KBARS, by the very nature of its construction, had a high degree of content validity. The degree of convergent validity was low, whereas the degree of discriminant validity was moderate. The assessment of convergent validity was based on the fact that there was a significant difference between the ratings by athletes and by the staff as a whole. It is possible that this significant difference among the groups was a function of the multirater method. Each set of ratings may have been valid for each specific rater group.

The data analysis indicated that there existed ten common elements or dimensions of coaching performance. The KBARS was judged to be a valid and reliable instrument for assessing these ten elements of the performance of the ten head coaches in the women's athletic program at Michigan State University during the 1977-1978 academic year.

### Conclusions

Feasibility of BARS for Assessing Coaching Performance. Few investigators have used the factor analytic method to study the

results of BARS although this is the recommended method (Smith, 1976). Landy and others (1976) used an eight-and-nine dimensional BARS for the appraisal of police officers by peers and supervisors. From the data of both BARS, they extracted three factors which accounted for approximately 78% of the variance in an oblique rotation. The intraclass coefficients ranged from .47 to .73, while the scale reliability coefficients ranged from .58 to .68. Landy and Guion (1970) factor analyzed a seven-dimensional BARS created to measure the work motivation of engineers. An orthogonal varimax rotation enabled them to extract four factors which accounted for 85% of the variance. Millard (1975) constructed a BARS with six dimensions for use by supervisors in rating their subordinates in the Department of Labor. He was able to extract one factor for each dimension.

Obviously, the KBARS results were unique in that only one factor could be extracted from the ten dimensions which accounted for all of the variance. Attempts to extract more than one factor were unsuccessful. This factor illustrated the contention by Smith (1976) that job traits are not orthogonal. The fact that just one factor could be extracted may have been due in part to the degree of communality shown by the Personality dimension with all the other dimensions. Perhaps for the population of raters used in the study, the coaches' personalities affected their performance in the other nine dimensions. On the other hand, this dimension could also have been a main source of leniency error. In other words, if the

coach was well-liked, s/he was rated effective in the other dimensions regardless of the anchors. Judging by the interdimension correlations, it might be feasible to exclude the personality dimension from the KBARS since it shared much variance with the other dimensions. However, the high loading of this dimension on the KBARS factor indicated that the coach's personality characteristics formed a strong part of this factor. It is the investigator's opinion that this dimension should not be excluded from subsequent evaluations.

Since there were no other existing data with which to compare the KBARS data, the extent of the superiority, if any, of the KBARS as compared to the data generated by numerical or graphic scales is unknown. However, its rigorous development and the data from Part III indicated that the BARS procedures were applicable to the appraisal of coaches at Michigan State University in 1977-1978.

The low convergent validity and high leniency of the KBARS indicated that the ratings were not error free. Such errors, however, may have been a function of the type of appraisal attempted, rather than a drawback of the instrument itself. Dickinson and Tice (1977) stated that a BARS could show high convergent validity with resulting ratings being incorrect inferences and vice versa. Borman (1978) attempted to construct and administer a BARS under an "ideal" situation and concluded that a certain amount of error will always be part of the results. He stated that:



utilizing a multitrait-multirater analysis to determine the convergent and discriminant validity seems unreasonable for organizational situations in which raters from different levels: (a) have significantly different orientations toward the job in question, and/or (b) observe significantly different ratee behavior. In other words, if raters from different levels have different relationships with ratees and see different occurrences of ratee behavior due to each level's unique vantage point, a criterion of agreement in ratings between these two groups seems forced at best (1974:106).

It is possible that the two different rater groups viewed each coach from different levels and depicted the situation described by Borman. The self-ratings showed the greatest amount of leniency which confirmed the results of others who have investigated the multirater method with rating scales in higher education and physical education. Centra (1972) found that both the self and peer ratings of college instructors showed strong leniency. Sweeting (1972) compared the ratings of high school physical education teachers by students, administrators and the teachers, themselves. He found that the latter set of ratings differed significantly from those of the other groups. These investigators also pointed out that leniency effects were not necessarily illustrative of error, but of different perspectives. What was critical behavior for one group was not necessarily critical for another.

The KBARS procedures could, however, generate less error. During the retranslation procedure each item was assigned a numerical value and allocated to dimensions. Perhaps there also should have been an assessment as to the frequency of such behaviors. This would allow for the elimination of all items which did not illustrate

"typical" behaviors. The ensuing scales would be anchored by behaviors which are "typical." Possibly the current KBARS contained some behavioral anchors which failed to illustrate common actions by coaches at Michigan State University. This may have required the raters to assign a rating by a subjective inference process and thus increase rater error.

Although the KBARS procedures did not yield error free results and ratings, the instrument has value beyond the actual scores and profiles. In itself each dimension is an informal statement of performance goals and a detailed description of appropriate job behaviors (Zedeck & Blood, 1974). This facet may have been its greatest asset since those who look at coaching performance tend to evaluate it in global and/or trait oriented terms, instead of specific behaviors. This was illustrated by the nature of the incidents initially collected and the comments generated by Part III of KBARS. Both Bernardin (1977) and Blood (1974) agreed that such "spin offs" from BARS procedures are just as important as the resultant ratings.

Implications of the KBARS Results for Coaching. Since the subjects involved in this investigation were not selected randomly, it is difficult to discuss the implications of the KBARS results for the coaching profession as a whole. However, some institutions which have similar athletic programs could examine the results and consider their applicability to their programs.

A behaviorally anchored instrument is vital to the assessment of coaches at Michigan State University as was evidenced by the type

of incidents generated by Phase I of the procedures and Part III of the KBARS. Many of the "incidents" were global attributes instead of specific behavioral examples. This tendency was also noted in the literature concerning coaching performance and was discussed in Chapter II. Most of the beliefs cited were abstract concepts, whereas the only concrete guidelines for assessment were those dealing with the won/lost record. Thus, most coaches do not have specific behavioral guidelines to follow and such an instrument is sorely needed.

The KBARS results have other implications for those in the women's athletic department at Michigan State University. It was found that the coach's personality was highly related to the raters' perceptions of the other nine dimensions. This echoed Percival's (1971) conclusion that the coach's personality may be the ultimate limitation to coaching success. KBARS, however, did not indicate the type of personality needed to guarantee success, but did state behavioral guidelines for the coaches to follow. The behaviorally anchored items gave an indication as to what a coach has to do, rather than what she has to be to experience coaching success at Michigan State University.

There was a significant difference between the scores of the ratings by the athletes and those of the staff. It was not clear if this divergence was due to an inadequacy of the KBARS or to an inherently different perspective of the coach by the two groups. A similar lack of convergence has been found in ratings of college

instructors by students and staff (Centra, 1972; Doyle, 1975). Regardless of the reason for the divergence, those involved in the women's athletic program at Michigan State University should be aware of this phenomenon. Use of the multirater method for the appraisal of coaches should be continued. In addition, the coach should be aware that how s/he and the staff perceived her/his performance was significantly different from the perceptions of the athletes. Reasons for this difference should be explored and where possible, eliminated.

### Conclusions

Ten different dimensions of coaching performance were observed in the women's athletic program at Michigan State University during 1977-1978 with the use of a behaviorally anchored rating scale. The quality of coaching was high as the ratings and comments illustrated. The 10 dimensions were highly related to each other, so that the KBARS functioned as a unity. The resultant KBARS showed a high degree of reliability and content validity, moderate discriminant validity, low convergent validity, and evidence of halo effect as well as leniency error. Until a more accurate instrument can be constructed, the KBARS was judged to be a viable method of evaluating the performance of coaches in this setting. Since there were differences between the ratings of different groups, the multirater method should be continued until the accuracy of each set of ratings can be investigated.

### Recommendations for the KBARS

1. All the dimensions of the current KBARS should be retained until a more accurate instrument can be developed. The accuracy of the results, however, could be increased if one person, possibly an administrative assistant, is put in charge of administering the KBARS and orienting the raters to common rating errors. This person would administer the KBARS to every team at the end of each season which should increase the return rate, reduce some of the response set bias, and ensure that the orientation of the raters to the evaluation is similar.

2. Part I of the instrument should be eliminated. It was impossible to determine to what degree it decreased the halo effect. At times it caused a great deal of confusion. Some of the raters could not understand its inclusion since it seemed to demand an assessment of the coach on the same dimensions as the KBARS and thus was a duplication and waste of time.

3. Part III of the instrument should be retained. The cited incidents should be reviewed for updating the KBARS. The women's athletic administrator found the comments generated by this section helpful in counselling and giving feedback to the staff. The only drawback to the inclusion of Part III is that it lengthens the time required to complete the instrument.

4. Since the retranslation process consisted of a procedure similar to factor analysis, perhaps the dimensions could be further refined by the actual use of factor analysis during this phase. In

other words, the retranslation sample would still assign numerical values to each statement, but those that met the standard deviation criterion would be subject to a factor analysis. Dickinson and Tice (1977) used this procedure and compared the results with those of a traditionally prepared BARS. They found that the factors of the former instrument were clearer and that the results were less prone to error. The main obstacle to the use of this method is based on the fact that Rummel (1970) and Aleamoni (1976) have stated that if the results of a factor analytic process are to be accurate, there should be minimally as many cases as there are variables. In the current study, 299 items underwent the retranslation procedure. The population that used KBARS numbered 173 and thus, even if the entire population were involved in the retranslation process, this rule of thumb would be violated. At this point also the number of items which would have failed to meet the standard deviation criterion if it has been applied first is unknown. Perhaps if it had been the first criterion instead of the dimension allocation criterion, more items might have been lost because of their variability than were lost in the current study. If the resultant number of items meets the criterion for factor analysis in conjunction with the size of the retranslation sample, then factor analysis should be a part of the retranslation procedure.

5. A method to reduce the time needed by the sample for the retranslation procedure should be created. It may be possible that

if a factor analysis were a part of this phase that a reduction in time may be the result since this would require the sample members to assign values to the items without having to allocate items to dimensions. However, if the typical frequency of occurrence of each behavior has to be assessed as was recommended, then the retranslation process will still be an arduous one. Perhaps the tasks could be spread over several smaller random samples (Harari & Zedeck, 1974). Selection procedures for both sample membership and item inclusion would have to be very rigorous so that all the data from the different samples could be combined.

7. The multirater method should continue to be employed, although there was a significant difference in the ratings from several groups. Since there is no criterion measure available to indicate which set of ratings were most accurate, it should be assumed that all raters rated as accurately as possible from their own perspectives.

8. The KBARS should be updated periodically using items from Part III of the instrument. Greater stress should be placed on providing examples of behaviors, rather than just comments. Future modification should include a concerted effort to generate incidents which would reduce the mid-range gap in the scales and those which apply to the Public Relations/Recruiting dimension.

9. The possibility of allowing the raters to mark each behavioral anchor as to whether it is relevant to the behavior of the coach being appraised should be investigated. Such a procedure,

if feasible, may indicate to what extent the leniency effect is a function of rater error or quality coaching.

#### Suggestions for Further Study

1. The KBARS should be administered to those in the men's athletic program at Michigan State University. If the results are reliable and valid, they should indicate to what extent the philosophies of the two existing athletic programs are congruent. Although one of the properties of the KBARS is its specificity, it is not known at this time to what extent the behavioral anchors are a function of sex and/or program.

2. The results of the KBARS ratings during the 1977-1978 season should be compared with those of the 1978-1979 season to determine the extent of the stability of the KBARS and the changes in coaching behaviors, if any.

3. Other Big Ten institutions or those with similar programs should develop BARS for the appraisal of coaches so that normative data can be established.

4. The results of the KBARS should be compared to data from other variables pertaining to the coaching situation, such as won/lost record, years of coaching experience, and team size. Such a comparison may give an indication as to which variables are most closely related to coaching performance as measured by the KBARS.



## BIBLIOGRAPHY

- AIAW Handbook 1977-1978. 1977. Washington, D.C.: AAHPER Publications.
- Albaugh, G. R. 1972. The influence of resentment as identified in college basketball coaches. Proceedings of the National College Physical Education Association for Men, 1972, pp. 60-68.
- Aleamoni, L. M. 1976. The relation of sample size to the number of variables in using factor analytic techniques. Education and Psychological Measurement, 36: 879-883.
- Andrud, W. E. 1970. The personality of high school, college and professional football coaches as measured by the Guilford-Zimmerman Temperament survey. Unpublished master's thesis, University of North Dakota.
- Bailey, C. 1972. Athletes' perception of coaches. Unpublished doctoral dissertation, University of Toledo.
- Bernardin, H. J. 1977. Behavioral expectation scales versus summated scales: A fairer comparison. Journal of Applied Psychology, 62: 422-427.
- Bernardin, H. J. 1978. Effects of rater training on leniency and halo errors in the ratings of instructors. Journal of Applied Psychology, 63: 301-308.
- Bernardin, H. J., Alvares, K. M., & Cranny, C. J. 1976. A recomparison of behavioral expectation scales to summated scales. Journal of Applied Psychology, 61: 564-570.
- Bernardin, H. J., LaShells, M. B., Smith, P. C., & Alvares, K. M. 1976. Behavioral expectation scales: Effects of developmental procedures and formats. Journal of Applied Psychology, 61: 75-79.
- Bernardin, H. J. & Walter, C. S. 1977. Effects of rater training and diary keeping on psychometric errors in rating. Journal of Applied Psychology, 62: 64-69.
- Blood, M. 1974. Spinoffs from behavioral expectation scale procedures. Journal of Applied Psychology, 59: 513-515.

- Borman, W. C. 1974. The ratings of individuals in organizations: an alternate approach. Organizational Behavior and Human Performance, 12: 105-124.
- Borman, W. C. 1978. Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63: 135-144.
- Borman, W. C. & Dunnette, M. D. 1975. Behavior-based versus trait-oriented performance ratings: An empirical study. Journal of Applied Psychology, 60: 561-565.
- Borman, W. C. & Vallon, W. R. 1974. A view of what can happen when behavioral expectation scales are developed in one setting and used in another. Journal of Applied Psychology, 59: 197-201.
- Burnaska, R. F., & Hollmann, T. D. 1974. An empirical comparison of the relative effects of rater response biases on three rating scale formats. Journal of Applied Psychology: 59: 307-312.
- Butts, D. S. 1976. Psychology of sport. New York: Van Nostrand Reinhold, Co.
- Campbell, D. T., & Fiske, D. W. 1959. Convergent and discriminant validity by the multitrait-multimethod matrix. Psychological Bulletin, 56: 81-105.
- Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. V. 1973. The development and evaluation of a behaviorally based rating scale. Journal of Applied Psychology, 57: 15-22.
- Campion, J. E., Greener, J., & Wernli, S. 1973. Work observation versus recall in developing behavioral examples for rating scales. Journal of Applied Psychology, 58: 286-288.
- 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.
- Cascio, W. F., & Valenzi, E. R. 1978. Relations among criteria of police performance. Journal of Applied Psychology, 63: 22-28.
- Centra, J. A. 1972. The utility of student ratings for instructional improvement. Princeton, N.J.: Educational Testing Service.

- Clark, M. J. 1975. An assessment of characteristics of successful women intercollegiate athletic coaches. Unpublished doctoral dissertation, Indiana University.
- Clay, J. T. 1974. Personality traits of female intercollegiate athletes and coaches. Unpublished doctoral dissertation, University of Utah.
- Counsilman, J. 1968. The science of swimming. Englewood Cliffs, N.J.: Prentice-Hall Inc.
- Cratty, B. J. 1973. Psychology in contemporary sport: Guidelines for coaches and athletes. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Danielson, R., Zelhart, P., & Drake, C. J. 1975. Multidimensional scaling and factor analysis of coaching as perceived by high school hockey players. Research Quarterly, 46: 323-334.
- Deatherage, D., & Reid, C. P. 1977. Administration of women's competitive sports. Dubuque, Iowa: Wm. C. Brown Co. Pub.
- DeCotiis, T. A. 1974. The development and evaluation of behaviorally anchored rating scales for the job of police patrolman. Unpublished doctoral dissertation, University of Wisconsin-Madison.
- DeCotiis, T. A. 1977. An analysis of the external validity and applied relevance of three rating formats, Organizational Behavior and Human Performance. 19: 247-266.
- Dickinson, T. L. & Tice, T. E. 1973. A multitrait-multimethod analysis of scales developed by retranslation, Organizational Behavior and Human Performance. 9: 421-438.
- Dickinson, T. L., & Tice, T. E. 1977. The discriminant validity of scales developed by retranslation, Personnel Psychology. 30: 217-228.
- Doyle, K. O. 1975. Student evaluation of instruction. Lexington, Mass.: Lexington Books, D. C. Heath & Co.
- Dressel, P. L. and Associates. 1961. Evaluation in higher education. Boston: Houghton Mifflin Co.
- Dunnette, M. D. 1966. Personnel selection and placement. Belmont, Cal.: Wadsworth Pub. Co.

- Dunnette, M. D. 1976. Aptitudes, abilities and skills. In Dunnette, M. D. (ed.) Handbook of industrial and organizational psychology. Chicago: Rand McNally Pub. Co.
- Ebel, R. L. 1951. Estimation of the reliability of ratings, Psychometrika. 16: 407-424.
- Eble, K., ed. 1970. The recognition and evaluation of teaching. Salt Lake City: Project to Improve College Teaching.
- Eble, K. 1972. Professors as teachers. San Francisco: Jossey-Bass, Inc., Pub.
- Edwards, H. 1973. Sociology of sport. Homewood, Ill.: Dorsey Press.
- Emery, D. B. 1962. Scoring the factors in successful high school coaching, Scholastic Coach. 31: 60, 62, 64-65.
- Flanagan, J. C. 1954. The critical incident technique, Psychological Bulletin. 51: 327-355.
- Fogli, L., Hulin, C. L., & Blood, M. R. 1971. Development of first level behavioral job criteria. Journal of Applied Psychology, 55: 3-8.
- Friedman, B. A., & Cornelius, E. T. 1976. Effect of rater participation in scale construction on the psychometric characteristics of two rating scale formats. Journal of Applied Psychology, 61: 210-216.
- Friedrich, J. A. 1953. Evaluating the coach. Athletic Journal, 33: 42-45, January.
- Frost, R. 1971. The director and his staff. In Steitz, E. (ed.) Administration of athletics in colleges and universities. Washington, D. C.: AAHPER, pp. 241-267.
- Gaintner, G. L. 1976. Athletes' and coaches' expectations of coaches. Unpublished doctoral dissertation, Arizona State University.
- Gallon, A. J. 1974. Coaching: Ideas and ideals. Boston: Houghton Mifflin Co.
- Goodale, J. C., & Burke, R. J. 1975. Behaviorally based rating scales need not be job specific. Journal of Applied Psychology, 60: 389-391.

- Harari, O., & Zedeck, S. 1974. Development of behaviorally anchored scales for the evaluation of faculty teaching. Journal of Applied Psychology, 58: 261-265.
- Hendry, L. B. 1969. A personality study of highly successful and ideal coaches. Research Quarterly, 40: 229-304.
- Hendry, L. B. 1972. The coaching stereotype. In Whiting, H. (ed.) Readings in sport psychology. Lafayette, Ind.: Balt Pub., 24-54.
- Hendry, L. B. 1974. Coaches and teachers of physical education: A comparison of the personality dimensions underlying their social orientation. International Journal of Sport Psychology, 5: 40-59.
- Hunt, G. 1974. Interscholastic and intercollegiate athletics for girls and women: Where do we go from here? In Proceedings: Women in sport. Provocative ideas for new directions. University of Rochester, May 15-18, pp. 63-69.
- Jones, D. 1975. Critical requirements for coaching women's intercollegiate volleyball in the state of Illinois. Unpublished master's thesis, University of Western Illinois.
- Jordan, J. B. 1976. The development and evaluation of behaviorally anchored rating scales for the assessment of department chairman performance. Unpublished doctoral dissertation, University of Kentucky.
- Kafry, D., Zedeck, S., & Jacobs, R. 1976. The scalability of behavioral expectation scales as a function of developmental criteria. Journal of Applied Psychology, 61: 519-522.
- Kavanagh, M. J., MacKinney, A., & Wolins, L. 1971. Issues in managerial performance: Multitrait-multimethod analysis of ratings. Psychological Bulletin, 75: 34-49.
- Keaveny, T. J., & McGann, A. F. 1975. A comparison of behavioral expectation scales and graphic rating scales. Journal of Applied Psychology, 60: 695-703.
- Kerlinger, F. N. 1973. Foundations of behavioral research. 2nd ed. New York: Holt, Rinehart and Winston, Inc.
- Klafs, C. E. & Lyon, M. J. 1977. The female athlete. 2nd ed. St. Louis: C. V. Mosby Co.

- LaGrand, L. E. 1970. A semantic differential analysis of the behavioral characteristics of coaches. Unpublished doctoral dissertation, Florida State University.
- Landy, F. J., Farr, J. L., Saal, F. E., & Freytag, W. R. 1976. Behaviorally anchored scales for rating the performance of police officers. Journal of Applied Psychology, 61: 750-758.
- Landy, F. J., & Guion, R. M. 1970. Development of scales for the measurement of work motivation. Organizational Behavior and Human Performance, 5: 93-103.
- Latham, G. P., & Wexley, K. N. 1977. Behavioral observation scales for performance appraisal purposes. Personnel Psychology, 30: 255-230.
- Lawler, E. 1967. The multitrait-multirater approach to measuring managerial job performance. Journal of Applied Psychology, 51: 369-381.
- Lawlis, G. F., & Lu, E. 1972. Judgment of counseling process: Reliability, agreement and error. Psychological Bulletin, 78: 17-20.
- Lawther, J. D. 1977. Psychology of coaching. 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Ley, K. 1973. Philosophical considerations in the administration of intercollegiate athletics for women. In Hunsicker, P. (ed.) Administration theory and practice in athletics and physical education. Chicago: Athletic Institute, pp.43-53.
- Loy, J. 1969. Social psychological characteristics of innovators. American Sociology Review, 34: 80-82.
- Maclean, N., & Wilner, B. 1977. Soviet sports exercise programs. New York: Drake Publishers Inc.
- Massengale, J. 1974. Coaching as an occupational subculture. Phi Deltan Kappa, 59: 140-142.
- McKinney, W. C. 1970. What is a good coach? Scholastic Coach, 39: 86.
- Meyer, D. W. 1972. A comparison of perceptions of university basketball coaches. Unpublished doctoral dissertation, University of Utah.

- Millard, C. 1975. The development and application of behavioral criteria for evaluating manpower performance at the micro-level. In Bedeian, A. G. (ed.) Proceedings: Academy of management. Thirty-fifth annual meeting. New Orleans, pp. 219-221.
- Miller, D. M. 1974. Coaching the female athlete. Philadelphia: Lea & Febiger.
- Morrow, J. R. 1976. The effects of response position and item simplicity on the factorial structure of teacher evaluation forms. Unpublished doctoral dissertation, University of Colorado at Boulder.
- Motowidlo, S. J., & Borman, W. C. 1977. Behaviorally anchored scales for measuring morale in military units. Journal of Applied Psychology, 62: 117-183.
- Neal, P. 1969. Coaching methods for women. Reading, Mass.: Addison Wesley Pub. Co.
- Neal, P. & Tutko, T. A. 1975. Coaching girls and women: Psychological perspectives. Boston: Allyn and Bacon Inc.
- Norton, S. D., Gustafson, D. P., & Foster, C. E. 1975. Assessment of management potential: Scale design, training effects and rater/ratee sex effects. In Bedeian, A. G. (ed.) Proceedings: Academy of management. Thirty-fifth annual meeting. New Orleans, pp. 237-239.
- Nunnally, J. C. 1967. Psychometric theory. New York: McGraw-Hill.
- Ogilvie, B. C., & Tutko, T. A. 1966. Problem athletes and how to handle them. London: Pelham Books.
- Ogilvie, B. C., & Tutko, T. A. 1970. Self-perception as compared with measured personality of selected male educators. In Kenyon, G. S. (ed.) Contemporary psychology of sport. Chicago: Athletic Institute, pp. 73-77.
- Palmieri, E. 1974. Evaluating the coach. The Athletic Director, 5: 4-5.
- Patrow, R. 1971. Psychosocial characteristics of coaches and their relationship to coaching success. Unpublished doctoral dissertation, University of Utah.
- Paul, G. 1977. Seven commandments. Grand Rapids Press. July 25, p. 3c.

- Percival, L. 1971. The coach from the athlete's viewpoint. In Percival, L. (ed.) International symposium on the art and science of coaching. Vol. I. Willowdale, Ontario: F. I. Productions, pp. 285-297.
- Peters, D. L., & McCormick, E. J. 1966. Comparative reliability of numerically anchored versus job-task anchored rating scales. Journal of Applied Psychology, 50: 92-96.
- Pflug, J. 1974. Evaluating coaches. The Athletic Director, 5: 2-3.
- Poindexter, H. 1973. Women's athletics: Issues and directions. In Proceedings: Annual Meeting, National Physical Education Association for Men, Kansas City, pp. 58-63.
- Poindexter, H. B., & Mushier, C. L. 1973. Coaching competitive sports for girls and women. Philadelphia: W. B. Saunders Co.
- Purdy, R. L. 1973. The successful high school athletic program. West Nyack, N.Y.: Parker Pub. Co. Inc.
- Ralbovsky, M. 1974. Lords of the locker room. New York: Wyden Pub. Co.
- Ralston, J., White, M., & Wilson, S. 1975. What is coaching? In Massengale, J.D. (ed.) The Principles and Problems of Coaching. Springfield, Ill.: Charles C. Thomas, pp. 5-9.
- Resick, M. C., & Erickson, C. E. 1975. Intercollegiate and interscholastic athletics for men and women. Reading, Mass.: Addison-Wesley Pub. Co.
- Rice, H. M. 1959. Qualities of a good coach. Bulletin of the National Association of Secondary Principals, 43:
- Rohlf, F. J., & Sokal, R. R. 1969. Statistical tables. San Francisco: W. H. Freeman and Company.
- Rummel, R. J. 1970. Applied factor analysis. Evanston, Ill.: Northwestern University Press.
- Sabock, R. 1973. The coach. Philadelphia: W. B. Saunders Co.
- Safrit, M. J. (ed.) 1976. Reliability theory. Washington, D.C.: AAHPER Publications.



- Sage, G. 1974a. The coach as management: Organization leadership in American sport. In Sage, G. (ed.) Sport and American society. 2nd ed. Reading, Mass.: Addison-Wesley Pub. Co., 429-439.
- Sage, G. 1974b. Machiavellianism among college and high school coaches. In Sage, G. (ed.) Sport and American society. 2nd ed. Reading, Mass.: Addison-Wesley Pub. Co., 187-206.
- Sage, G. 1974c. Value orientations of American college coaches compared to those of male college students and businessmen. In Sage, G. (ed.) Sport and American society. 2nd ed. Reading, Mass.: Addison-Wesley Pub. Co., 207-228.
- Sage, G. 1975. An occupational analysis of the college coach. In Ball, D. W. & Loy, J. W. Sport and the social order: contributions to the sociology of sport. Reading, Mass.: Addison-Wesley Pub. Co., 359-455.
- Schneider, B. 1976. Staffing organizations. Pacific Palisades, Cal.: Goodyear Pub. Co.
- Schwab, D. P., Heneman, H., & DeCotiis, T. A. 1975. Behaviorally anchored rating scales: A review of the literature. In Bedeian, G. (ed.) Proceedings: Academy of management. Thirty-fifth annual meeting, New Orleans, pp. 222-224.
- Shirley, J. D. 1966. Profile of an ideal coach. Journal of Health, Physical Education and Recreation, 37: 59.
- Siegel, S. 1956. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.
- Singer, R. N. 1972. Coaching, athletics, and psychology. New York: McGraw-Hill.
- Slater, S. J. 1976. The development of rating scales for the purpose of controlling leniency bias in student evaluations of college instructors. Unpublished doctoral dissertation, University of California at Santa Barbara.
- Smith, P. 1976. Behaviors, results, and organizational effectiveness: The problem of criteria. In Dunnette, M. D. (ed.) Handbook of industrial and organizational psychology. Chicago: Rand McNally Pub. Co.
- Smith, P., & Kendall, L. M. 1963. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47: 149-155.

- Smith, R. E., Smoll, F. L., & Hunt, E. 1977. A system for the behavioral assessment of coaches. Research Quarterly, 48: 401-407.
- Sokal, R. R., & Rohlf, F. J. 1969. Biometry. San Francisco: W. H. Freeman and Co.
- Sweeting, R. L. 1972. Perceived personality characteristics and effectiveness of physical education teachers. Unpublished doctoral dissertation, University of Illinois.
- Tharp, R. G., & Gallimore, R. 1976. What a coach can teach a teacher. Psychology Today, 10: 75-78, January.
- Underwood, J. 1962. The desperate coach. Sports Illustrated, 31: August 25: 66-76; September 1: 20-27; September 8: 28-40.
- Women's Athletic Handbook. 1978. Unpublished manuscript. East Lansing, Michigan: Women's Athletics, Michigan State University.
- Zedeck, S., & Baker, H. 1972. Nursing performance as measured by a behavioral expectation scale: A multitrait-multirater analysis. Organizational behavior and human performance, 7: 457-466.
- Zedeck, S., & Blood, M. R. 1974. Foundations of Behavioral Science Research in Organizations. Monterey, Cal.: Brooks/Cole Pub. Co.
- Zedeck, S., Jacobs, R., & Kafry, D. 1976. Behavioral expectations: Development of parallel forms and analysis of scale assumptions. Journal of Applied Psychology, 61: 112-115.
- Zedeck, S., Kafry, D., & Jacobs, R. 1976. Format and scoring variations in behavioral expectation evaluations. Organizational Behavior and Human Performance, 17: 171-184.

APPENDIX A

LETTER TO ATHLETES

LETTER TO COACHES

BEHAVIORAL STATEMENTS - PHASE I

EXAMPLES OF BEHAVIORAL STATEMENTS

August 29, 1977

Dear Athlete:

As part of an attempt to continually upgrade the women's athletic program at Michigan State University, a comprehensive coaching appraisal form will be developed during the 1977-1978 year. This appraisal form needs your input so that it can accurately reflect coaching performance. Many rating scales consist of categories such as "strongly agree" or "good, average, fair." These words can be interpreted in many different ways and thus, the resultant ratings can vary widely. Consequently this appraisal form will consist of a rating scale which is anchored on statements of observable behaviors. Since the rating of each coach will be done by his/her team, administrator, other coaches, the coach him/herself and support personnel, it is essential that these people are involved in scale construction. This makes the rating scale applicable to all coaches.

Your input is needed in two ways:

First of all, please list 3-5 effective behaviors and 3-5 ineffective behaviors on the accompanying sheet. Each of these should be one sentence which describes an action by a coach in women's sports, which you have observed. This sheet should be postmarked no later than September 16.

The second phase of the scale's development will involve, among others, three athletes from each team. To ensure that the scale is statistically sound, these athletes will be chosen randomly (i.e. picked out of a hat). If you are selected for Phase II, you would attend one evening session (Monday, October 3, 7:00 - 9:30 p.m.), in which you would assign numerical values to behavioral statements and assign the statements to categories. The resultant data will be used to finalize the coaching evaluation scale and will be administered at the end of each sports season.

Since the development of such a rating scale is an elaborate and detailed procedure and since it must be ready for use by fall sport teams, it is essential that Phase I and II go as smoothly as possible. You can help by returning the behavioral statement sheet on time (or earlier). We need your input even though you may have decided not to participate this year. If you are selected for Phase II, you will receive a letter which will finalize the details.

It seems that whenever a rating scale is produced, those who do the rating grumble and complain because it doesn't meet their criteria. This is your chance to ensure that the ratings of coaches in women's athletics at MSU are accurate and interpretable, as well as reflecting the differences between sports.

Sincerely,

Annelies Knoppers  
Athletic Staff

August 29, 1977

Dear Coach:

As part of an attempt to continually upgrade the women's athletic program at Michigan State University, a comprehensive coaching appraisal form will be developed during the 1977-1978 year. This appraisal form needs your input so that it can accurately reflect coaching performance. Many rating scales consist of categories such as "strongly agree" or "good, average, fair." These words can be interpreted in many different ways and thus, the resultant ratings can vary widely. Consequently this appraisal form will consist of a rating scale which is anchored on statements of observable behaviors. Since the rating of each coach will be done by his/her team, administrator, other coaches, the coach him/herself and support personnel, it is essential that these people are involved in scale construction. This makes the rating scale applicable to all coaches.

Your input is needed in two ways:

First of all, please list 3-5 effective behaviors and 3-5 ineffective behaviors on the accompanying sheet. Each of these should be one sentence which describes an action by a coach in women sports, which you have observed. This sheet should be postmarked no later than September 16.

Please encourage all of your athletes to submit such statements. Winter and spring sport athletes will have received a letter in the mail similar to this one. If you are a fall sport coach, I would like an opportunity to meet with your team at its first meeting if possible, so that they can submit such statements, as soon as possible. Please leave a note in my box or call me as to when your first meeting will be held.

Phase II of the development of the rating scale will involve two athletes from your team as well as yourself. Three returning athletes will be selected randomly from your team. They will be asked to assign numerical values to behavioral statements and to assign statements to categories. It will involve one evening (Monday, October 3, 7:00 - 9:30 p.m.). To make the system as statistically sound as possible, it is imperative that those chosen to participate in Phase II do so. You will be notified as to which athletes have been selected. (They will be notified by letter.) Please take some time to encourage them to attend. You will also be asked to participate in Phase II, but will be able to do so on your own time, since it is easier to track you down than it is students! Your

cooperation and support is very much appreciated. The quality of the rating system will be proportional to the amount of input from all of those involved in women's athletics.

To summarize, I am asking you for the following:

1. To write a list of statements which describe effective-ineffective behaviors of coaches--postmarked September 16 or in my box by September 19.
2. To encourage your athletes to send in their statements--winter and spring sports.
3. To give me ten to fifteen minutes of time at your initial team meeting--fall sports.
4. To make sure your three selected athletes attend the Monday, October 3 meeting.
5. To participate in Phase II and to complete that form as soon as possible when it becomes available.

The resultant data will be used to construct a rating scale. Each coach will be rated by her/his athletes, the other coaches, assistant coach, trainer, Nell, Cheryl, managers, intern (if any) and her/himself at the end of a season. With your support and input such a rating will be an accurate profile of your coaching performance in behavioral terms. If you have any questions, please feel free to contact me.

Sincerely,

Annelies Knoppers

## Behavioral Statements - Phase I

**Introduction:** The general purpose of the women's athletic program is to help each player and team develop potential. Think of times when a coach did something you considered to be very effective or very ineffective. These incidents should "stand out" in your mind in order to be considered. Such incidents do not necessarily have to involve you and a coach, but should be incidents you have observed. The emphasis is on observable behaviors not qualities.

**Instructions:**

1. In each space give a statement which is descriptive of a behavior you have observed.
2. In each case, state the statement as illustrating an effective (E) or ineffective (I) behavior.
3. These behavioral statements can involve any coach and any sport, but must be observable.
4. List a minimum of three in each category (effective/ineffective). If you can list more, do so.
5. Put only 1 statement in each space.
6. Return the sheets, so they are postmarked no later than September 16.

**Example:** E - The coach was usually at practice by the time the first player arrived.

I - The coach yelled so loud at the team during a timeout that the opposing team could hear her.



Behavioral Statements

Your Sport(s) at MSU \_\_\_\_\_ Are you a MSU:

Athlete \_\_\_\_\_  
Coach \_\_\_\_\_  
Trainer \_\_\_\_\_  
Other \_\_\_\_\_

1. E = Effective I = Ineffective

2. E or I?

3. E or I?

4.

5.

6.

7.

8.

9.

## EXAMPLES OF BEHAVIORAL STATEMENTS

Effective

By being organized on away trips, the coach allowed the girls to concentrate on their competition.

The coach shows an interest at practice and competition.

The coach's directions are short and to the point.

The coach demonstrates superior knowledge of the sport.

The coach established ground rules by which athletes operate and adhere to rules.

The coach was concerned about all of us and when something was wrong, tried to help.

The coach offered to help players after practice on our own time.

The coach showed an interest and voiced encouragement at practice and competition and so the athlete's outlook and efforts took on a positive and more hopeful attitude.

The coach was always at practice.

Ineffective

The athlete received no feedback from the coach.

The coach approached competition with many negative comments about the athlete's ability to do well.

The coach depended on one person to be the whole team.

The coach joked around and acted like one of us kids too much.

The coach had unpredictable moods--one day friendly, the next day, mad at the world.

The player made a mental error and the coach chewed her out in front of the remaining players.

The coach physically abused the players.

The coach turned his back on the team (literally) and went up and sat in the bleachers during the game.

Practices were long and boring.

Disagreements between coach and assistant coach were not solved before crucial contests.

APPENDIX B

COACHES EVALUATION PROJECT--PHASE II

## Coaches Evaluation Project--Phase II

## Directions:

On the following page are listed twelve dimensions related to different aspects of coaching performance at Michigan State University. These dimensions or aspects have been identified on the basis of input by athletes, trainers, coaches and athletic administrators involved in the women's athletic program at Michigan State University. The twelve dimensions are lettered in alphabetical sequence A through L.

Also included in this packet are 299 statements which describe examples of coaching behaviors observed by athletes and athletic personnel. Your job is to make two judgments concerning each behavioral statement:

1. Dimension Assignment--After reading the statement, please state the letter (A through L) which best typifies the dimension with which the specific statement of behavior should be identified. I would suggest that you detach the dimension descriptions and refer to them when assigning examples to dimensions. You may only chose one dimension per statement and must chose one. Even though some statements may seem similar, try to assign each one independently as much as possible.
2. Item Rating--After assigning a statement to a dimension, rate the statement with respect to the quality of performance it represents by assigning to it a number from 1-7. A value of 1 represents extremely ineffective performance; 3-4 represent average coaching performance and 7 represents extremely effective performance. Try to assign ratings in terms of general (typical) performance of coaches in the overall program, rather than just in terms of your specific sport. Use only whole numbers and rate each statement independently.

Work as fast or slowly as you wish, but try to complete the project at one time so that you are in the same frame of mind. Please turn in all materials, write legibly and work independently. The results will be incorporated in an instrument used to evaluate coaching performance on the basis of observed behaviors. Your time and care in completing this project helps to insure that the instrument is valid and reflects the values of those involved in the women's athletics program at Michigan State University. Thank you for your cooperation, time, and effort.

## DIMENSIONS

- A. Behaviors concerned with scheduling and organizing home and away contests. (Scheduling)\*
- b. Behaviors concerned with the organization and context of practices (Practice)
- C. Behaviors concerned with team selection and coaching philosophy. (Selection)
- D. Behaviors which illustrated personal facets of the coach. (Personality)
- E. Behaviors concerned with the setting of team rules, discipline and team control. (Discipline)
- F. Behaviors concerned with skill development in practice and the use of these skills in contests. (Skill)
- G. Behaviors concerned with the performance of the athletes in actual competitive experiences. (Meets)
- H. Behaviors which affected the athlete psychologically and provided motivation and support for her. (Support)
- I. Behaviors related to communication with the team and with promotion of teamwork. (Communication)
- J. Behaviors which denoted professional characteristics. (Professionalism)
- K. Behaviors which denoted a knowledge of the sport and coaching techniques. (Knowledge)
- L. Behaviors concerned with public relations and recruiting. (P.R./Recruiting)

\*The bracketed name indicates the code name which was used for the dimension throughout the study. The code names were not part of this instrument.

## Behavioral Statements

- \*1. The coach talks openly on the bench about an athlete who is playing poorly while other athletes are within hearing distance (Personality)<sup>a</sup>
- \*2. The coach incorporates a new drill about every other practice (Practice)
- 3. When a player makes a mental error, the coach chews her out in front of the other athletes (Meets)
- 4. The coach has favorites (Personality)
- 5. The coach spends an equal amount of time with each athlete during the week (Meets)
- 6. The coach gives constructive criticism (Meets)
- 7. Practices are well organized (Practice)
- \*8. During tough times the coach encourages rather than complains (Support)
- \*9. In the first practice of the season the coach sits down with the athletes and asks each one to talk about herself (Communication)
- \*10. Home meets/matches are well organized (Scheduling)
- \*11. The coach makes team rules (i.e. curfew) and always enforces them (Discipline)
- \*\*12. The coach participates (plays) with the athletes in scrimmages and does not watch to see what the players are doing wrong (Personality)
- \*\*13. The coach offers to help individuals after practice (Support)
- 14. The coach depends on one athlete to be the whole team (Selection)
- 15. When something is wrong with one of the athletes, the coach can tell and tries to help (Meets)

<sup>a</sup>The bracketed dimension(s) represent(s) the dimension(s) with the modal frequency.

\*Item included in KBARS

\*\*Item retained but not included in KBARS

- \*16. The coach jokes around and acts like one of the kids (Personality)
- \*17. The coach benches the best athlete because she has failed to do the conditioning program (Discipline)
- \*18. The coach gets excited when the team does well (Personality)
19. The coach does not get upset if athletes "goof off" in practice (Skill)
20. The coach does not know the sport well enough (Practice/Knowledge)
21. The coach is unsure of herself and transmits the feelings to the team (Meets)
22. The coach helps with personal problems (Support)
23. The coach is able to get the team "up" for all matches/meets (Support)
- \*24. The coach lets players skip practices often (Discipline)
25. The coach is at practice before it starts (Practice)
26. The coach insists everyone be well stretched out before the workout begins (Knowledge)
- \*27. The coach teaches the most effective techniques to the athletes (Knowledge)
28. The coach is rarely critical (Personality)
29. The coach corrects every single mistake every single time (Skill)
30. The coach's directions are usually short and to the point (Professionalism/Knowledge)
31. Disagreements between coach and assistant coach are not resolved before a crucial contest (Meets)
32. The coach frequently checks with the trainer as to the physical status of the athletes (Professionalism)
- \*33. Athletes arrive late for practice with no subsequent action taken by the coach (Discipline)



- \*34. The coach praises an area of performance before giving suggestions to improve (Support)
- 35. The coach varies partners or players within groups (Practices/Communication)
- \*36. Athletes are always involved in drills and are rarely standing around (Practice)
- 37. The coach leads a one-sided discussion (give and no take) at team meetings (Meets)
- 38. The coach ignores an athlete's lack of hustle (Meets)
- 39. The coach frequently emphasizes the importance of the star athlete (Communication)
- \*40. The coach frequently cancels practice (Practice)
- \*41. The athletes are so intent on pleasing the coach, they lose track of the reason they are in the sport (Selection)
- 42. The coach approaches competition with many negative comments about the ability of her athletes (Support)
- \*\*43. The athlete rarely receives feedback from the coach. (Support)
- \*44. The coach is organized on away trips (Scheduling)
- 45. The coach expresses despair when the opponents look stronger than her team (Personality/Support)
- 46. The coach builds up team morale when the team is in a difficult situation (Support)
- 47. The coach treats all athletes the same regardless of skill and ability (Skill)
- \*48. The coach openly expresses her disgust with the team while they're competing (Support)
- \*49. No specific instructions are given during time outs (Knowledge)
- 50. The coach is generally not perceptive of feelings and attitudes of her team (Communication)
- \*51. The coach does not share the itinerary of an away trip with her team far enough in advance (Schedule)

52. The coach gives a rationale for each drill (Knowledge)
- \*\*53. The team is properly conditioned for the type of activity and level of competition (Knowledge)
54. The coach takes into account individual differences in personality of her athletes (Support)
- \*55. The coach gives thorough explanations (Knowledge)
56. The coach talks in low whispers to the assistant coach at various times in practice (Practice)
57. During team meetings the coach listens but does not hear what is being said (Meets)
- \*58. The coach shows very little emotion during and after contests (Personality)
59. The coach tells the athletes what is expected of them (Selection)
- \*60. The coach does not always give immediate feedback on error correction (Knowledge)
61. If a player does not work hard, the coach makes her run laps (Discipline)
- \*62. The coach puts up a "suggestion" envelope expecting all team members to give input (Communication)
63. The coach does conditioning exercises/workouts with the team (Personality/Communication)
- \*\*64. The coach runs a drill in which team members do not use proper technique because it seems useless (Knowledge)
- \*\*65. The coach helps individuals on a one to one basis (Support)
66. The coach listens when you talk and then gives her reasons explaining what she does and why (Professionalism)
67. The coach expects hustle from the athletes when they are fatigued, and exhausted (Meet)
68. An athlete is disciplined for not trying even though there could be a reason for it (Discipline)
69. The coach always takes time to talk with her athletes (Communication)

70. The coach listens to what the players say, but will not always change her mind or adjust (Personality/Communication)
71. The coach is open to suggestions and incorporates them (Communication)
72. The coach states the goals of a drill in reference to the game/meet situation (Skill)
73. The coach encourages the use of the trainer (Professionalism)
74. The coach uses patronizing psychology (Meets)
75. The coach lends moral support when the team is fatigued and ready to give up (Support)
76. The coach makes athletes swim extra laps for being late when the lateness was not their fault (Discipline)
- \*77. The coach changes drills when things slow down rather than yelling at the players for moving so slowly (Knowledge)
- \*78. The coach starts and pulls players without saying anything to them (Support)
- \*79. During practice a coach pats a player on the back after the athlete has just completed an extremely difficult drill (Support)
- \*\*80. The coach points out that the athlete has used the correct technique immediately after the athlete performs it (Knowledge)
81. The coach does not get upset at a physical error as long as she knows the athlete is trying her hardest (Support)
82. The coach frequently yawns during practice (Personality)
83. The coach allows athletes who become bogged down with school work to miss an occasional practice (Discipline)
84. The coach ignores an athlete's bad attitude (Discipline)
85. The coach makes an effort to have a close relationship with her athletes (Personality)
86. The coach never yells at her athletes (Personality)
- \*87. The coach has a tendency to leave the athletes in the dark about situations involving the team (Communication)

88. The coach is very high on upper classmen helping freshmen (Communication)
- \*\*89. The coach can verbalize her knowledge of skill mechanics (Knowledge)
- \*90. The coach can laugh when appropriate--has a sense of humor (Personality)
91. The coach shows favoritism when choosing a starting team (Personality)
- \*92. The coach is inconsistent with praise (Support)
- \*93. The coach's moods are unpredictable (Personality)
94. The coach can be approached for personal problems (Personality)
95. The coach frequently arrives late for practice because she is so busy (Skill)
96. The coach brings her infant to the gym during workouts (Professionalism)
97. The coach attempts to keep contact with each member of the team during the off season and summer (Communication)
98. The coach sets up an individual training schedule for the off season for those athletes who request it (Skill)
- \*99. The coach gives many helpful tips during practices (Knowledge)
- \*100. The coach takes the time to encourage team members before and after practices (Support)
101. The coach posts the list for trips the night before departure so that all team members have had an equal chance to show they they should go (Selection)
- \*\*102. The coach says to her team during a game timeout "I don't know what else to say to you" (Knowledge)
103. The coach likes to leave immediately after a match, and, thus, the athletes do not have time to socialize with the opponents (Personality)
104. The coach tells the second string that they are as good as the starters, but that the line has to be drawn somewhere (Meets)

- \*105. The coach tells the team that a certain strategy will be ineffective against the opponents, then in overtime, she makes the team use that strategy (Knowledge)
- \*106. When a player asks the coach what is wrong with her technique the coach says, "I don't know, ask. . . ." (Knowledge)
- 107. The coach does not allow her athletes to socialize with the opponents (Personality)
- 108. The coach stresses winning ahead of individual achievement and progress (Selection)
- 109. The coach seems to disregard the fact that some of her athletes are in pain (Personality/Professionalism)
- 110. The coach likes to see the team suffer (Personality)
- 111. The coach comes to every practice (Professionalism)
- 112. The coach always criticizes and never gives credit for accomplishment (Support)
- \*113. The coach keeps up with the newest ideas in his/her sport (Knowledge)
- 114. The coach spends most of her time with those players who need the most help (Skill)
- \*115. The coach remembers a new athlete from previous correspondence with that athlete (P.R./ Recruiting)
- 116. The coach uses statistics as well as observations to make judgments concerning the team (Knowledge)
- 117. The coach is flexible (Personality)
- \*118. The coach follows department and university policies and rules (Professionalism)
- \*119. The coach always discusses the decisions of the officials with them (Professionalism)
- 120. The coach tries to draw the officials' attention to fouls made by the opponents (Professionalism)
- 121. The coach emphasizes the fact that each athlete is responsible for herself (Discipline/Support)

- \*122. The coach always congratulates the opposing team's coach (Professionalism)
- \*123. The coach has more knowledge of the game than any other person on the team (Knowledge)
- \*124. The coach gives each person an equal opportunity to demonstrate her talent (Selection)
- \*125. The coach adheres to rules and regulations which she sets up for her athletes (Discipline)
- 126. The coach is strict (Discipline)
- \*127. The coach is often insulting (Personality)
- 128. The coach and athletes do not relate well off the court (Personality)
- \*129. The coach is prejudiced and a bigot (Personality)
- \*130. The team rules are fair (Discipline)
- 131. The team is always in excellent physical condition (Meets)
- \*132. The coach's team rules are unrealistic (Discipline)
- 133. The coach uses a variety of techniques to motivate the team (Support)
- 134. The coach is unwilling to make sacrifices in order to help the team become successful (Personality)
- 135. The coach is a symbol of maturity (Personality)
- 136. The coach is unable to remove herself from the role of an active participant (Personality)
- 137. The coach uses only ethical means of motivation (Professionalism)
- 138. The coach emphasizes the values of competitive athletics (Selection)
- 139. The coach stresses proper dietary and health measures to her athletes (Discipline)
- \*\*140. The team's schedule of opponents is challenging (Schedule)

141. The coach is extremely serious about developing a good team (Selection)
142. The coach makes an athlete feel proud in front of other teams that she is their coach (Professionalism)
- \*143. The coach works very closely with the captain (Communication)
- \*144. The coach praises the whole team as well as individuals (Communication)
145. The coach is aware of individual differences and, therefore, expects more of some athletes than others (Selection/Support)
- \*146. The coach helps the team get to know everyone (Communication)
147. The varsity rarely scrimmages the junior varsity (Skill)
- \*148. Team members receive very little individual help (Support)
149. The coach always gives statistical as well as verbal feedback the day after a game/meet (Knowledge)
- \*150. Practices hardly vary (Practice)
151. The coach's enthusiasm for the sport is obvious (Personality)
- \*152. The team travels a great deal (Schedule)
- \*\*153. Praise is given for a fine performance (Support)
154. The coach discusses her personal problems with team members (Personality)
- \*155. The coach often dwells on the financial situation of the team (Professionalism)
156. The coach does little to get the team mentally prepared for a meet/match (Support)
157. The coach reveals the lineup a day prior to the meet (Selection)
158. The coach makes the athletes work extra hard the day after tournaments or championships (Knowledge)
- \*159. The coach makes remarks which make an athlete cry (Support)

- \*\*160. The coach always has the practices/workouts planned (Practice)
- 161. The coach expects too much from the athletes (Support)
- \*162. News about what is going on is not given out to the entire team (Communication)
- 163. The coach overworks the athletes so that some drop out (Selection)
- 164. The coach expects most of the team spirit to come from the athletes (Personality/Communication)
- 165. After a "close-but-not-close-enough" game the coach does not say anything to a player (Support)
- 166. The coach tries to deal with problems and conflicts instead of ignoring them (Communication)
- 167. The coach takes time to publicly recognize his/her staff (Meets)
- 168. The coach is concerned about the athlete's schoolwork (Professionalism)
- \*169. The coach criticizes the athlete even though the athlete feels she has done the best she can (Support)
- \*170. The coach shows little confidence in the team's ability to win an important event giving the impression that no matter how hard they try they cannot win (Support)
- \*\*171. The coach gives personalized individual attention to athletes whenever possible (Support)
- 172. The coach yells at a player during a timeout in front of a field house full of people (Support)
- 173. The coach tries to maintain the idea that in the end you should enjoy your sport; although it should be taken seriously, it should never be "all work and no play" (Selection)
- 174. The coach has the attitude that the sport she is coaching is the only sport worth participating in (Personality)
- 175. The coach uses a great deal of physical and emotional closeness in helping the athlete (Support)
- \*\*176. The coach criticizes each shot the athlete makes (Support)



177. Coaches from other schools take advantage of the coach and get away with it (Professionalism)
178. The coach often asks team members if they want to work on something special (Communication)
179. The coach asks that a match be delayed an extra hour so a team member can make the tournament because of a class conflict (Schedule)
180. The coach always takes care of any school problems the players have (Professionalism)
181. Although the coach is at all practices she is often not attentive (Personality)
- \*182. The coach keeps records of each individual's performance in workouts/practices (Practices)
- \*183. The coach rarely carries through her threats (Discipline)
184. The coach gets angry at an athlete and then ignores her (Personality)
185. During an away meet, the coach usually insists that all athletes stick together rather than allowing them to go off with their friends (Discipline)
- \*186. The coach talks critically about the team to athletes in other sports (Professionalism)
- \*\*187. The coach is openly critical of another coach at MSU in front of athletes (Professionalism)
- \*188. The coach is openly critical of another coach at another school in front of her team (Professionalism)
189. The coach shows patience and faith in each individual (Meets)
190. The coach says that this year will be a building year because not many players are back from last year (Support)
191. The coach discusses with an athlete on the team the faults of a team member (Professionalism)
- \*192. The coach was very successful last year so she is using the same strategies she used then even though she has totally different athletes (Knowledge)

- \*193. The coach does not stress weight training (Knowledge)
- \*194. The content of workouts/practices are always written on the board (Practice)
- \*195. The coach suggests various exercises to help the individual athlete (Knowledge)
- \*196. The coach delivers criticism of technique privately (Support)
- \*197. The coach uses drills designed for men (Knowledge)
- 198. The coach pays no attention while the junior varsity competes (Meets)
- 199. The coach rarely offers information concerning related area in women's athletes (P.R., Recruiting)
- \*\*200. The coach gives little instruction between events/ games (Knowledge)
- 201. The coach slacks down on workouts under pressure of the athletes (Practice)
- \*202. The coach suspends team members for missing practices (Discipline)
- \*203. The coach often stays overtime to help athletes with specific techniques (Support)
- \*204. The coach refuses to allow athletes to be interviewed by the press (P.R., Recruiting)
- 205. The coach lacks social manners and dresses inappropriately during competition which embarasses team members (Meets)
- 206. The coach frequently compares the inferiority of the women's sports program to that of the men at MSU (Skill)
- \*207. The coach frequently complains of communication problems between herself and the administration (Professionalism)
- 208. The coach frequently complains about his teaching load (Professionalism)
- 209. The coach explains the "why" in workouts (Practice)
- 210. The coach tells her athletes she will pull anyone from the game who makes a mental error (Support)

211. The coach continually stresses the fact that the team is defending champion (Support)
212. The coach does not always start the best players (Selection)
- \*213. The coach rarely has the team scrimmage in practice (Practice)
- \*214. The coach increases the intensity of the workouts 3 days before a championship meet (Knowledge)
- \*215. The coach does not switch drills until everyone does the drill well (Practice)
216. The coach explains that she is not ready for practice because she has had a bad day (Personality)
217. The coach continually pushes the concept of "pride" (Support)
- \*218. The coach constantly re-evaluates athletes' performance and will change line-ups and entries accordingly (Selection)
- \*219. The coach does things which are not the most effective for the individual but do contribute to the team as a whole (Communication)
- \*\*220. The coach takes the time to talk to those athletes who are nervous before the meet (Support)
221. The coach constantly corrects the athlete's technique (Skill)
222. The coach hugs an athlete who has just done her personal best (Personality/Meets)
- \*\*223. The coach uses the videotape frequently for technique correction (Knowledge)
224. When an athlete is not one of the best performers, the coach does not pay much attention to her (Support)
225. The coach talks mainly to the returning athletes at the organizational meeting (Communication)
226. The coach puts up with athletes no one else can stand (Personality)
227. The coach allows the assistant coach to use threats, crude remarks and scare tactics to motivate athletes to better performance (Support)

228. When an athlete has an injury, the coach calls her at night to make sure she is ok. (Support)
- \*229. A coach pulls an injured athlete from an event/match even though that costs the team the win (Professionalism)
230. The coach catches the athlete's eye before an important event and smiles (Support)
- \*231. When an athlete is depressed about her performance the coach tells her she believes in her (Support)
232. The coach rarely gets down on the team for losing (Selection)
- \*\*233. The coach establishes her lineup just an hour before the meet begins (Selection)
- \*234. The coach takes the time to talk to a frustrated athlete (Support)
235. The coach yells at an injured player for not icing her injury when she was supposed to (Support)
- \*236. The coach displays much outward tension during meets (Personality)
- \*237. The coach becomes very emotional when the team wins a championship (Personality)
238. The coach really drives the athletes without any consideration for their feelings (Support)
239. The coach uses starters and second-string persons together throughout practice so that if in a game or relay situation a second stringer is called upon, she will feel as if she belongs (Knowledge)
240. The coach often gives athletes encouraging pats on the back in stressful situations (Support)
241. The coach uses drills which are related to performance in the upcoming championship (Skill)
242. The coach does not always follow her promises about letting some athletes start, substitute or compete (Selection)
243. The coach is never satisfied with anything but the best performance from his/her athletes (Personality/Meets)

- \*244. The coach strictly enforces curfews (Discipline)
- \*245. The coach posts the team list after a month of tryouts instead of making cuts earlier (Selection)
- 246. The coach does not allow her athletes to participate in intramural sports (Selection)
- 247. If an athlete does not make the traveling team she is ignored and not included in some of the team activities (Communication)
- 248. The coach tries to make the players independent of her (Personality/Professionalism)
- 249. The coach never praises or puts down to a strong degree, but instead by remaining neutral makes the athlete condemn or praise herself (Support)
- 250. The coach teaches the joy of being a competitor (Selection)
- \*251. The coach establishes a dress code for trips (Discipline)
- 252. The coach offers herself as a sort of counselor (Personality)
- 253. The coach always warm ups with the athletes (Personality/Communication)
- \*254. One drill always follows another so the team has no time to get bored (Practice)
- 255. The coach doesn't close her eyes to nonstarters (Selection)
- 256. The coach attends other sporting events (besides her own) at MSU (Professionalism)
- 257. The coach always implements the suggestions of the athletes even though she may not agree with them (Scheduling)
- 258. The coach has a party at her house at the beginning of the season, but will not let the players bring alcoholic drinks (Personality)
- \*259. In a close, but frustrating game, the coach says she can't think of anything else to do (Knowledge)
- 260. The coach ignores the team after they perform poorly in a match/meet (Support)

261. The coach controls the team mentally in a negative way--using scare tactics, etc. (Support)
262. The coach inadvertently leaves a name off the list of players who made the team (Selection)
- \*263. The coach does not allow her athletes to show anger or frustration in the game situation (Discipline)
- \*264. The coach pulls an athlete from a match even though the athlete was playing her heart out (Selection)
- \*265. The coach talks to all the athletes' parents whenever they attend a meet (P.R., Recruiting)
- \*266. The coach talks negatively about athletes and coaches in the other sports (Professionalism)
- \*267. The coach does not make concessions for tryouts for those athletes who are in overlapping sports (Selection)
268. The coach talks an athlete into concentrating on only one sport (Selection)
269. The coach demands more of scholarship athletes than non-scholarship athletes (Selection)
- \*270. The coach always seems to obtain the trip money and cars at the last minute (Schedule)
271. The coach uses the second string when the team is well ahead (Selection/Meets)
- \*272. The coach substitutes if an athlete is doing poorly (Selection)
- \*273. When the coach selects the team she keeps it as small as possible (Selection)
- \*274. The coach states his/her beliefs on the values of participation at the first team meeting (Selection)
- \*275. The coach uses instructional devices such as videotapes, dittos and charts (Knowledge)
- \*276. The coach frequently compliments the players (Support)
- \*277. Once the starting lineup has been established or athletes assigned to events, it tends to stay the same all season (Selection)

278. The coach stresses continued competition after the season (Selection)
279. The coach explains self and procedures by verbalizing her thoughts and actions (Communication)
280. The coach uses profane language occasionally (Personality)
281. The coach does not allow athletes to enter a contest if they exhibit lack of emotional control (Selection)
- \*282. The coach takes the joy out of playing because the practices are too hard (Practice)
283. The coach attends coaches' meetings irregularly (Professionalism)
284. The coach asks an athlete to play in a position (or compete in an event) where the athlete feels inadequate (Selection)
285. The coach says she does not care whether the team wins or loses as long as they play well (Selection)
286. The coach tells the athletes she is embarrassed by their performance at a contest (Support)
287. The coach stresses mental rehearsal and builds it into the warmup schedule (Skill)
288. The coach expects more of the veterans than the freshmen (Selection)
289. The coach emphasizes the fact that the team represents the university (P.R., Recruiting)
290. The coach asks the athletes how they would like to be coached (Selection)
291. The coach asks the athletes' help in deciding the team lineup (Selection)
292. The coach is good buddies with team members (Personality)
293. The coach lets the team know tactfully when she is disappointed in them (Communication)
294. The coach babies certain athletes (Support)
- \*295. The coach smokes (Personality)

- \*296. The coach is out of condition (Personality)
- 297. The coach puts himself wholly into coaching (Personality)
- 298. The coach selects uniforms and equipment for the team which are up-to-date and effective (Support)
- \*299. The coach handles correspondence in a reasonable period of time (P.R., Recruiting)



## Coaches' Evaluation Project - Phase II

## Answer Sheet

Name: \_\_\_\_\_ Telephone \_\_\_\_\_

Address: \_\_\_\_\_

Sport: \_\_\_\_\_ Circle one: \*Athlete \*Coach \*Assistant  
Coach  
\*Administration \*Trainer  
\*Other

Years involved in women's athletics at MSU \_\_\_\_\_

State- ment #	Dimen- sion A-L	Rating 1-7	Dimen- sion A-L	Rating 1-7	Dimen- sion A-L	Rating 1-7
1.	_____	_____	21.	_____	41.	_____
2.	_____	_____	22.	_____	42.	_____
3.	_____	_____	23.	_____	43.	_____
4.	_____	_____	24.	_____	44.	_____
5.	_____	_____	25.	_____	45.	_____
6.	_____	_____	26.	_____	46.	_____
7.	_____	_____	27.	_____	47.	_____
8.	_____	_____	28.	_____	48.	_____
9.	_____	_____	29.	_____	49.	_____
10.	_____	_____	30.	_____	50.	_____
11.	_____	_____	31.	_____	51.	_____
12.	_____	_____	32.	_____	52.	_____
13.	_____	_____	33.	_____	53.	_____
14.	_____	_____	34.	_____	54.	_____
15.	_____	_____	35.	_____	55.	_____
16.	_____	_____	36.	_____	56.	_____
17.	_____	_____	37.	_____	57.	_____
18.	_____	_____	38.	_____	58.	_____
19.	_____	_____	39.	_____	59.	_____
20.	_____	_____	40.	_____	60.	_____

APPENDIX C

COACHING EVALUATION PACKET (KBARS)

Michigan State University  
Women's Athletics

COACHING EVALUATION PACKET (KBARS)

You are asked to evaluate the coaching performance of a specific coach based on your perceptions of that coach's usual behaviors. It is important that you read instructions carefully so that the resultant ratings are an accurate and honest reflection of the coach. These ratings will be used to give the coach feedback as to your perceptions as well as guidelines for improvement for the future. Please take the time to complete the form carefully and accurately writing only on the answer sheet (not on this evaluation packet). Both the packet and answer sheet should be returned to 220 Jenison (Dr. Jackson) or to the person designated by the coach.

## Instructions - Part I

READ CAREFULLY

Below are listed ten categories/dimensions of coaching performance. For each one, rate your coach on the scales on the answer sheet. A value of one (1) represents extremely ineffective or very poor performance while a value of seven (7) represents extremely effective or very good performance. The middle values (3-4) indicate average performance. For each category put an X through the appropriate number on the answer sheet. Choose only one number per category. Do not write on this page. Complete Part I before going on to Part II.

## Categories

1. Scheduling and organization of home and away competition
2. Organization and content of practices
3. Team selection and coaching philosophy
4. Personal characteristics of the coach
5. Team rules, discipline and team control
6. Motivation and psychological support of athlete:relations with individual athletes
7. Communication with team and promotion of teamwork:relations with the team as a whole
8. Professionalism:relations with peers and support personnel
9. Knowledge of the sport and coaching techniques/strategies
10. Public relations and recruiting:relations with fans, parents, press and high school athletes and coaches

## Instructions - Part II

READ CAREFULLY - Part II

On the attached pages are scales which assess ten categories of coaching performance behavior. Each category is given on a single page. At the top of each page a dimension is briefly defined. Below the definition there is a seven-point rating scale incorporating behavioral statements. These are examples designed to help you define the scale values. You will need to judge as to which scale value best describes your coach. IT IS NOT NECESSARY THAT YOUR COACH HAS ACTUALLY DEMONSTRATED ANY OF THE BEHAVIORS LISTED. You are only making a comparison with what your coach could be expected to do.

1. Read the title and the behavioral examples on a page
2. For each category put an X through the number on the answer sheet (11-20), which corresponds to the scale value which best describes your coach on the average
3. Choose only one number per category, but read all examples before doing so
4. If more than one statement describes your coach, work with the statement which best describes his/her behavior on the average
5. DO NOT WRITE ON THE PAGES OF THE CATEGORIES

### 11. SCHEDULING AND ORGANIZATION OF HOME AND AWAY COMPETITION

This coach could be expected. . .

Extremely Ineffective Performance →

1-

-to obtain travel money and cars at the last minute

2-

-to share itineraries with his/her teams only at the last moment

3-

Average Performance →

4-

-to set up an extensive schedule so that the team travels a great deal

5-

6-

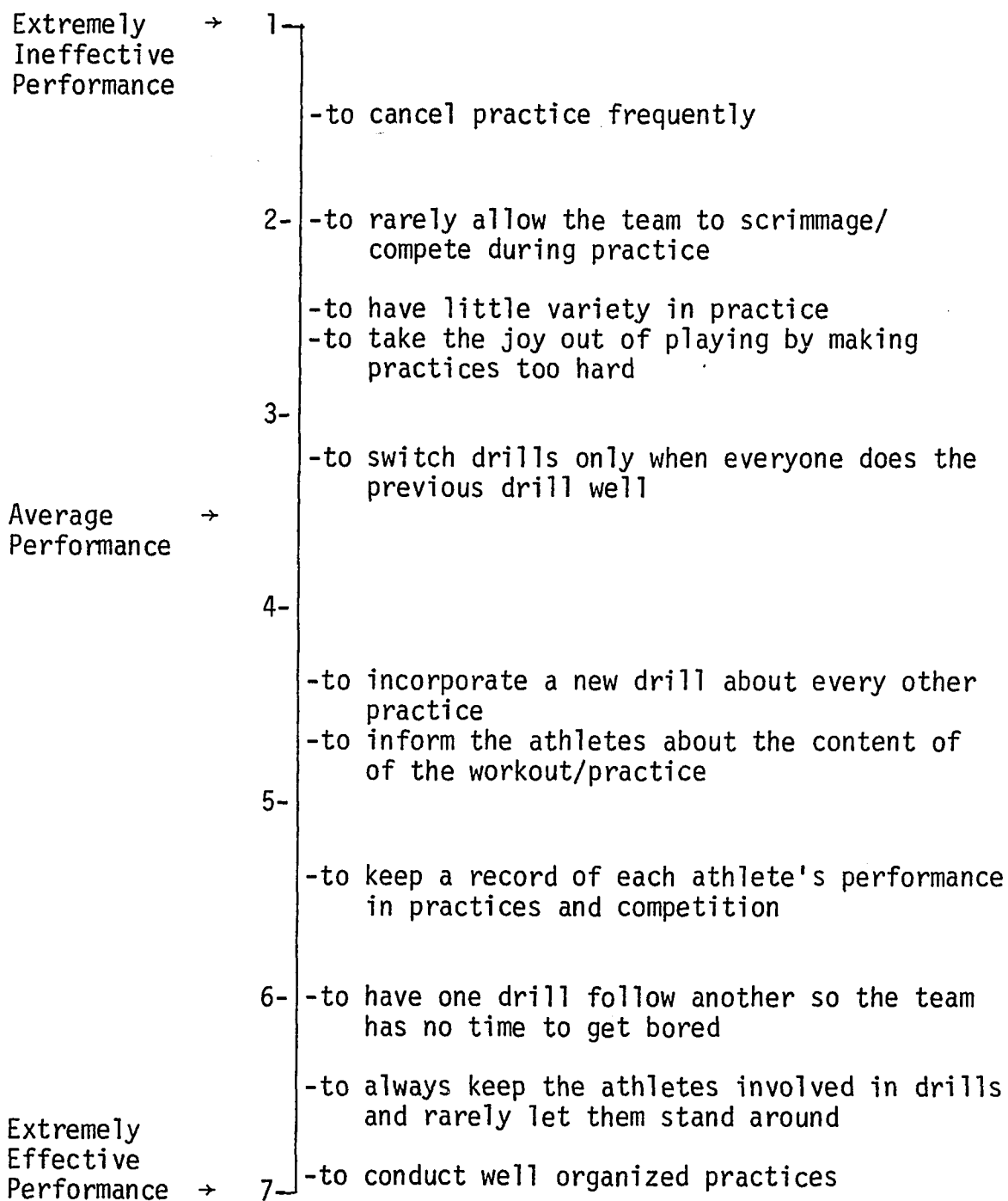
Extremely Effective Performance →

7-

-to be organized on away trips  
-to have home meets which are well organized

## 12. ORGANIZATION AND CONTENT OF PRACTICES

This coach could be expected. . .



## 13. TEAM SELECTION AND COACHING PHILOSOPHY

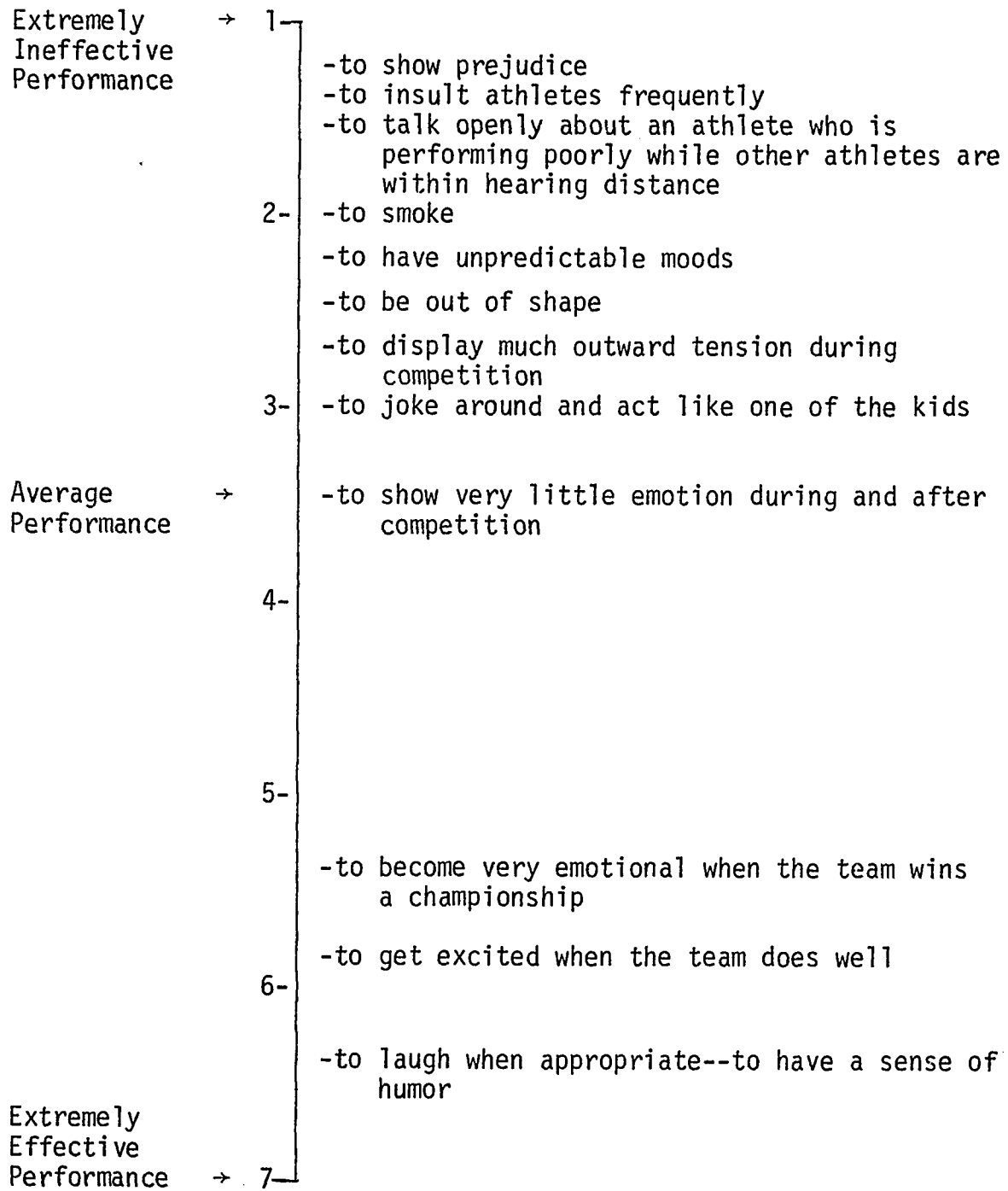
This coach could be expected. . .

Extremely Ineffective Performance	→ 1-	-to allow athletes to lose track of the reason they are in the sport by having them worry only about pleasing him/her
	2-	-to make no concessions for tryouts for those athletes who are in overlapping sports
	3-	-to keep the line-up or assignments to events the same all season
Average Performance	→	-to use at least a month for tryouts instead of making cuts earlier
	4-	-to pull an athlete from competition even though the athlete has been working her heart out
		-to select a team as small as possible
	5-	-to substitute if an athlete is doing poorly
		-to constantly re-evaluate each athlete's performance and change lineup/entries accordingly
	6-	-to state his/her philosophy of coaching at the team's first meeting
Extremely Effective Performance	→ 7-	-to give each athlete an equal opportunity to demonstrate her talent



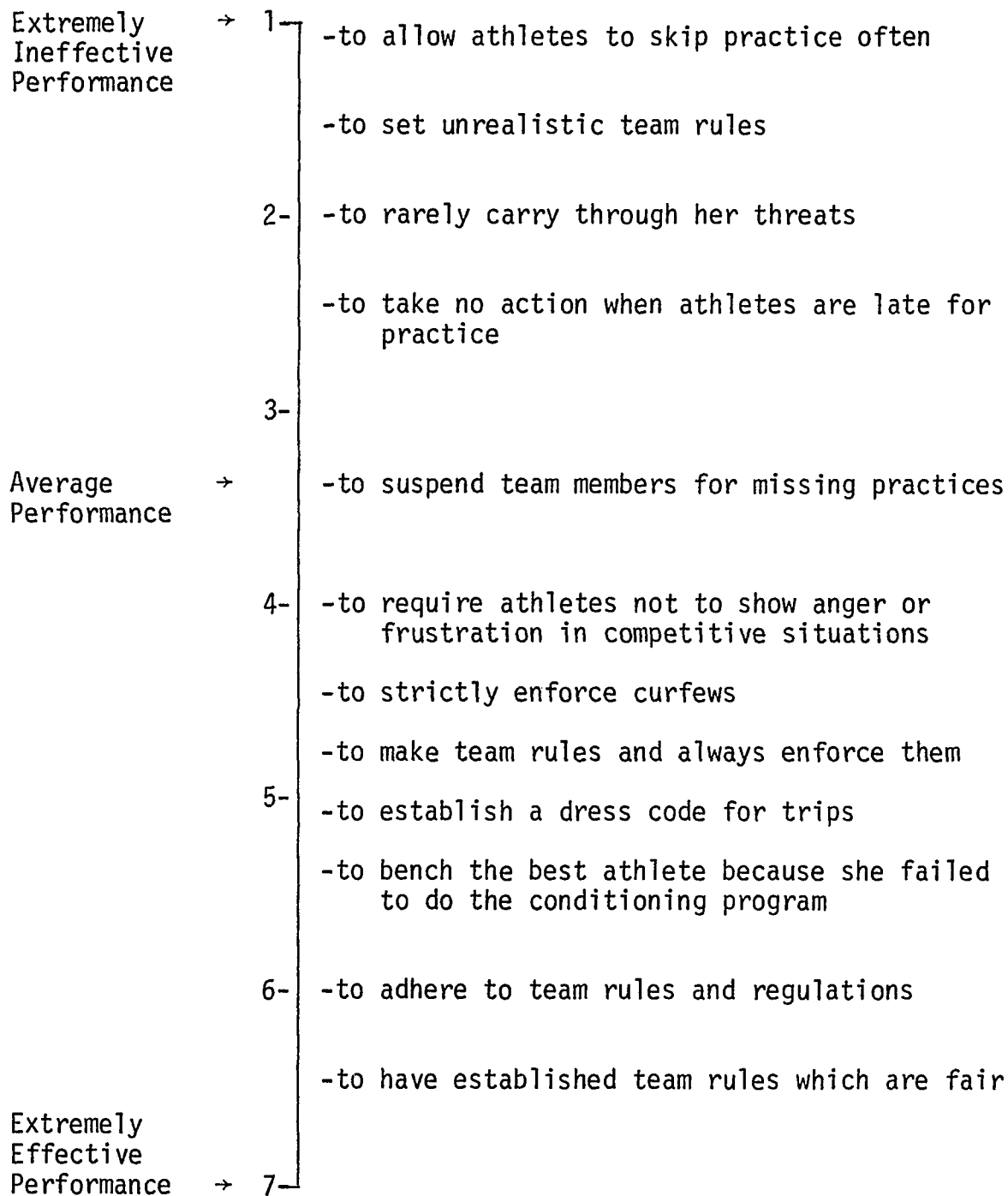
## 14. PERSONAL CHARACTERISTICS OF THE COACH

This coach could be expected. . .



## 15. TEAM RULES, DISCIPLINE AND TEAM CONTROL

This coach could be expected. . .



16. MOTIVATION AND PSYCHOLOGICAL SUPPORT OF ATHLETE:  
RELATIONS WITH INDIVIDUAL ATHLETES

This coach could be expected. . .

Extremely Ineffective Performance	→ 1-	<ul style="list-style-type: none"> <li>-to give the impression that no matter how hard the team tries they cannot win</li> <li>-to give athletes very little individual help</li> </ul>
	2-	<ul style="list-style-type: none"> <li>-to start and pull athletes without saying anything to them</li> <li>-to make remarks which would make an athlete cry</li> <li>-to openly show disgust with the team while it is competing</li> </ul>
	3-	<ul style="list-style-type: none"> <li>-to praise inconsistently</li> <li>-to criticize an athlete who thought she had done the best she could</li> </ul>
Average Performance	→	
	4-	
	5-	<ul style="list-style-type: none"> <li>-to praise an area of performance before giving suggestions for improvement</li> <li>-to criticize an athlete's technique privately</li> </ul>
	6-	<ul style="list-style-type: none"> <li>-to frequently compliment the athletes</li> <li>-to often stay overtime to help athletes with specific techniques</li> <li>-to give athletes encouraging pats on the back in stressful situations</li> <li>-to talk to a frustrated athlete</li> <li>-to take time to encourage team members before and after practice</li> <li>-to tell an athlete s/he believes in her when the athlete is depressed</li> </ul>
Extremely Effective Performance	→ 7-	<ul style="list-style-type: none"> <li>-to encourage rather than complain during tough times</li> </ul>

17. COMMUNICATION WITH TEAM AND PROMOTION OF TEAMWORK:  
RELATIONS WITH THE TEAM AS A WHOLE

This coach could be expected. . .

Extremely Ineffective Performance	→ 1-	-to talk about what is going on to only a few athletes
	2-	-to leave the athletes in the dark about situations involving the team
	3-	
Average Performance	→	
	4-	-to sit down with the team at the first practice and ask each athlete to talk about herself
		-to put up a "suggestion" envelope to obtain ideas from the athletes
	5-	-to do things which may not be the most effective for the individual athlete but which do contribute to the team as a whole
		-to work closely with the captain
	6-	-to help the team to get to know everyone
Extremely Effective Performance	→ 7-	-to praise the team as a whole as well as individuals

18. PROFESSIONALISM: RELATIONS WITH  
PEERS AND SUPPORT PERSONNEL

This coach could be expected. . .

Extremely Ineffective Performance	→ 1-	-to criticize his/her team to athletes in other sports
		-to talk negatively about athletes and coaches in other sports at MSU
	2-	-to frequently complain of communication problems with the administration
		-to openly criticize coaches in sports at other institutions
	3-	-to frequently complain about the financial situation of the team
Average Performance	→	-to always question the officials about their decisions
	4-	
	5-	
	6-	-to follow department and university policies
		-to pull an injured athlete from competition even though that action costs the team the win
Extremely Effective Performance	→ 7-	-to always congratulate the opposing coach

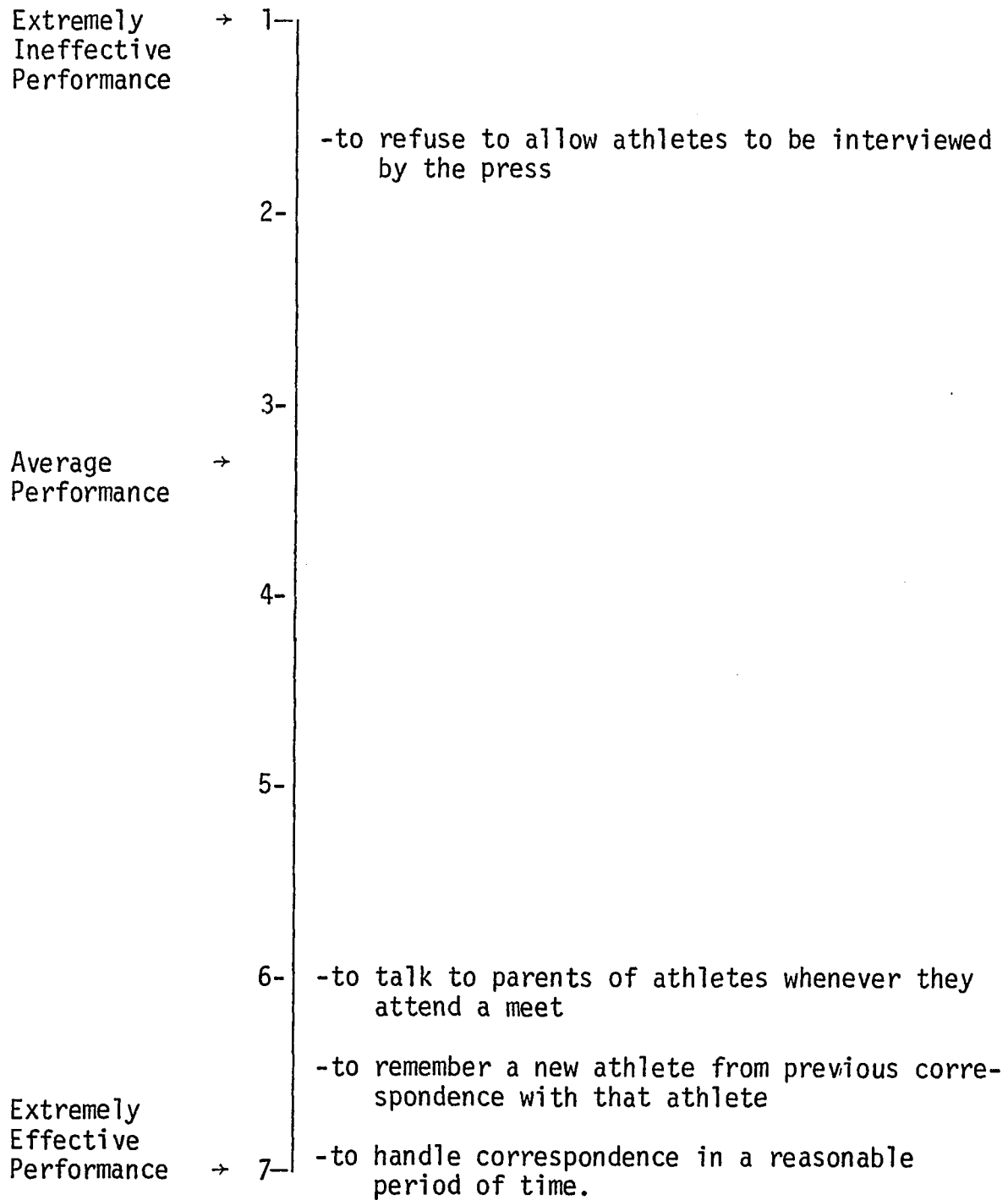
19. KNOWLEDGE OF THE SPORT AND COACHING  
TECHNIQUES/STRATEGIES

This coach could be expected to. . .

Extremely Ineffective Performance	→ 1-	-to tell a team s/he doesn't know what to do during a close but frustrating competitive event
		-to tell an athlete, who asks what is wrong with her technique, "I don't know, ask . . ."
	2-	-to fail to give specific instructions during timeouts
		-to tell the team that a certain strategy will be ineffective against the opponents and then use that strategy in overtime
		-to fail to stress weight training
	3-	-to use the same strategy this year as last year even though s/he has totally different athletes
Average Performance	→	-to increase the intensity of the workouts just before a championship meet
		-to fail to give immediate feedback on error correction
	4-	-to use drills designed for men
		-to change drills when things slow down rather than yelling at the athletes for moving so slow
	5-	-to give thorough explanations
		-to frequently use instructional devices such as videotape, dittos, and charts
		-to know more about the sport than any other team member
	6-	-to teach the most effective techniques
		-to suggest various exercises to help the individual athletes
		-to give many helpful tips during practice
Extremely Effective Performance	→ 7-	-to keep up with the newest ideas in his/her sport

20. PUBLIC RELATIONS AND RECRUITING: RELATIONS WITH FANS, PARENTS, PRESS, AND HIGH SCHOOL ATHLETES AND COACHES

This coach could be expected. . .



## Coaching Evaluation Answer Sheet

1. Are you: Coach (Self) Coach (Peer) Ass't Coach Athlete  
Trainer Other: \_\_\_\_\_
2. Last name of coach being rated \_\_\_\_\_
3. Sport \_\_\_\_\_
4. If you are an athlete: a. Fr. Soph. Jr. Sr.  
b. Varsity or Junior Varsity: \_\_\_\_\_  
c. Number of years on team at MSU (including this one  
and junior varsity) \_\_\_\_\_  
d. Number of years of eligibility left \_\_\_\_\_  
e. Are you team captain? Yes \_\_\_\_\_ No \_\_\_\_\_
5. If you are not an athlete: Number of years in women's  
athletics at MSU \_\_\_\_\_

Part I & II Put an X through the appropriate number. Note that the question numbers go across the page.

Part I: Make sure you rate every category!

- |                  |                   |
|------------------|-------------------|
| 1. 1 2 3 4 5 6 7 | 2. 1 2 3 4 5 6 7  |
| 3. 1 2 3 4 5 6 7 | 4. 1 2 3 4 5 6 7  |
| 5. 1 2 3 4 5 6 7 | 6. 1 2 3 4 5 6 7  |
| 7. 1 2 3 4 5 6 7 | 8. 1 2 3 4 5 6 7  |
| 9. 1 2 3 4 5 6 7 | 10. 1 2 3 4 5 6 7 |

Part II. Make sure you rate every category!

- |                   |                   |
|-------------------|-------------------|
| 11. 1 2 3 4 5 6 7 | 12. 1 2 3 4 5 6 7 |
| 13. 1 2 3 4 5 6 7 | 14. 1 2 3 4 5 6 7 |
| 15. 1 2 3 4 5 6 7 | 16. 1 2 3 4 5 6 7 |
| 17. 1 2 3 4 5 6 7 | 18. 1 2 3 4 5 6 7 |
| 19. 1 2 3 4 5 6 7 | 20. 1 2 3 4 5 6 7 |



Part III: Comment and/or give examples of behaviors which justify your ratings for Part I & II

1. Scheduling and organization of home and away competition
2. Organization and content of practices
3. Team selection and coaching philosophy
4. Personal characteristics of the coach
5. Team rules, discipline and team control
6. Motivation and psychological support of athlete: relations with individual athletes
7. Communication with team and promotion of teamwork: relations with the team as a whole
8. Professionalism: relations with peers and support personnel
9. Knowledge of the sport and coaching techniques/strategies
10. Public relations and recruiting; relations with fans, parents, press and high school athletes and coaches