INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
The effects of three grading systems on student outcome measures in table tennis

Mao, Youxiang, Ph.D.
The Ohio State University, 1991

Copyright ©1991 by Mao, Youxiang. All rights reserved.
THE EFFECTS OF THREE GRADING SYSTEMS ON STUDENT OUTCOME MEASURES IN TABLE TENNIS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree of Philosophy in the Graduate School of The Ohio State University

by

Youxiang Mao, B.A., M.Ed.

* * * * *

The Ohio State University

1991

Dissertation Committee:
Dr. D. Zakrajsek
Dr. D. Tannehill
Dr. D. Rosenberg

Approved by

Dr. Dorothy Zakrajsek
Advisor
School of Health, Physical Education, and Recreation
Copyright
by
Youxiang Mao
1991
To my Mother
For Your Giving as Sun to Grass
ACKNOWLEDGMENTS

My sincere thanks to my advisor, Dr. Zak, for your care and love, for your guidance and design of my coursework, for granting me teaching assistantship, and for the way you taught me to think and write precisely in preparing research.

To Dr. Tannehill, I thank you for your insights, help, guidance, and love in sharing the direction of my dissertation. You provided important "patriot missiles" in helping me win my war with finishing my dissertation.

To Dr. Rosenberg, I am grateful for your help, suggestions, and friendship during this last year.

To my peers, Jenny, Niki, and Tom, I thank you for helping me with coding training sessions. Special thanks to Jenny who helped me code lessons and provided warm tea.

To students in my classes, thank you. You were smart while sometimes sheep-like and cooperative yet independent, and I learned much from you.
VITA


1973-1978 .................. Basketball Player, GuiZhou, P.R. of China

1978-1983 .................. B.A., Department of Chinese Literature, SiChuang University, P.R. of China

1983-1985 .................. M.Ed., Department of Coaching, Beijing Institute of Physical Education.

1986-1987 .................. Assistant Professor, Department of Physical Education, GuiZhou Teacher's University, P.R. of China.

1987-1989 .................. Teaching Assistant, Division of Health, Physical Education, Recreation, and Dance, University of Idaho.

1989-1991 .................. Graduate Teaching Associate, School of Health, Physical Education, and Recreation, The Ohio State University.

PUBLICATIONS


FIELDS OF STUDY

Major Field: Health, Physical Education, and Recreation--Teacher Education in Physical Education
Professor
Dr. D. Zakrajsek
Dr. D. Tannehill (Co-Advisor)
Dr. D. Rosenberg

Minor Field: Applied Statistics
Professor
Dr. E. Novak
TABLE OF CONTENTS

DEDICATION .......................................... ii
ACKNOWLEDGEMENT ..................................... iii
VITA ................................................ iv
LIST OF TABLES ...................................... ix
LIST OF FIGURES ..................................... xi

CHAPTER PAGE

I. INTRODUCTION ........................................ 1

           Statement of the Problem ..................... 9
           Null Hypotheses ............................. 10
           Definitions .................................. 10
           Delimitations ............................... 12
           Limitations .................................. 13
           Significance of This Study .................. 14

II. THE REVIEW OF RELATED LITERATURE................. 16

           Important Assumptions and Theories
           Related to Grading ......................... 16
             Student Achievement as
                A Dependent Variable .................... 16
             Motivation Theory X ........................ 21
             Dunkin and Biddle Model .................... 23

           Research Paradigms Related to Grading ...... 24
             Process-Product Paradigm .................... 24
             Ecological Paradigm ......................... 28
             Experimental Teaching Units (ETU)
           in Research on Teaching in
             Physical Education ....................... 32

           Research on Grading in Physical
           Education .................................... 36
             Grading of Physical Education in
                Measurement/Evaluation and in

vi
## III. METHODOLOGY

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedures</td>
<td>59</td>
</tr>
<tr>
<td>Subjects</td>
<td>62</td>
</tr>
<tr>
<td>Design Model</td>
<td>63</td>
</tr>
<tr>
<td>Possible Threats to Internal Validity</td>
<td>64</td>
</tr>
<tr>
<td>Variation of Selection</td>
<td>64</td>
</tr>
<tr>
<td>Researcher Bias</td>
<td>65</td>
</tr>
<tr>
<td>Loss of Student Cooperation</td>
<td>66</td>
</tr>
<tr>
<td>Treatments or Independent Variables</td>
<td>66</td>
</tr>
<tr>
<td>Duration of Treatment</td>
<td>69</td>
</tr>
<tr>
<td>Dependent Variables</td>
<td>70</td>
</tr>
<tr>
<td>Student Behaviors</td>
<td>70</td>
</tr>
<tr>
<td>Attendance rates</td>
<td>71</td>
</tr>
<tr>
<td>One-Minute Volley Test</td>
<td>71</td>
</tr>
<tr>
<td>Serve Test</td>
<td>72</td>
</tr>
<tr>
<td>Written Examination</td>
<td>73</td>
</tr>
<tr>
<td>Skills Demonstrated on Game Play Variables</td>
<td>73</td>
</tr>
<tr>
<td>Game Play Format and Organization</td>
<td>74</td>
</tr>
<tr>
<td>Procedure for Game Play Recording</td>
<td>74</td>
</tr>
<tr>
<td>Reliability Check</td>
<td>75</td>
</tr>
<tr>
<td>Measures on Student Attitude Questionnaire</td>
<td>77</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>78</td>
</tr>
</tbody>
</table>

## IV. RESULTS

<table>
<thead>
<tr>
<th>Result</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Behavior Consistency Across the Three Treatment Classes</td>
<td>80</td>
</tr>
<tr>
<td>Comparisons of Grading Effects on Student Behaviors by DIBA Categories</td>
<td>84</td>
</tr>
<tr>
<td>Comparisons of Grading Effects on Student Attendance Rate</td>
<td>89</td>
</tr>
<tr>
<td>Comparison of Grading Effects on Skill Tests</td>
<td>91</td>
</tr>
<tr>
<td>Comparison of Grading Effects on</td>
<td>91</td>
</tr>
</tbody>
</table>
Written Test .......................... 93
Comparisons of Grading Effects on Game Play Variables ............... 96
Comparisons of Student Attitude Toward the Three Grading Systems ... 121

V. SUMMARY AND DISCUSSION, CONCLUSION, AND RECOMMENDATION .......... 128

Summary and Discussion ................. 128
Positive Effects of Grading Systems .... 129
Reversed Effects of Grading Systems Across Attitude and the Other Outcome Measures .................. 133
Non-Significant Measures on the Effects of Grading Systems ............ 137
The Contributions to Experimental Studies in Physical Education ........ 142
Cost of Measurement ................... 143
Difficulty of Measuring Learning Table Tennis Skills ................. 144

Conclusions ......................... 145

Recommendations for Future Research .... 148
Recommendations for Improving This Study .................. 148
Recommendations for Replicating This Study .................. 149
Recommendations for Grading Activity Classes in Physical Education .... 149

REFERENCES ......................................... 151

APPENDIXES ......................................... 161
A. Approval Form for Involvement of Human Subjects .................. 164
B. Course Outlines, Unit Teaching Plan, First Week Teaching Plan .... 173
C. Direct Instruction Behavior Analysis (DIBA) Coding Manual ........ 179
D. Written Test ................................ 188
E. Definition for Game Play Recording .................. 192
F. 15-Item Questionnaire on Attitude .................. 195
G. Mean Comparisons on Attitude Questions .................. 198
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Three grading systems used as independent variables</td>
<td>67</td>
</tr>
<tr>
<td>2. Grading systems for racket sports at OSU</td>
<td>69</td>
</tr>
<tr>
<td>3. DIBA data on teacher behaviors (Coder J.)</td>
<td>82</td>
</tr>
<tr>
<td>4. DIBA data on teacher behaviors (Coder M.)</td>
<td>83</td>
</tr>
<tr>
<td>5. Comparison of grading effects on student behavior by DIBA categories (Coder M.)</td>
<td>87</td>
</tr>
<tr>
<td>6. Comparison of grading effects on student percent of attendance rate</td>
<td>90</td>
</tr>
<tr>
<td>7. ANCOVA on one-minute volley tests</td>
<td>94</td>
</tr>
<tr>
<td>8. Mean comparison on one-minute volley test scores</td>
<td>94</td>
</tr>
<tr>
<td>9. ANCOVA on serve tests</td>
<td>95</td>
</tr>
<tr>
<td>10. Mean comparison on serve tests</td>
<td>95</td>
</tr>
<tr>
<td>11. ANCOVA on written scores</td>
<td>97</td>
</tr>
<tr>
<td>12. Mean comparison on written scores</td>
<td>97</td>
</tr>
<tr>
<td>13. ANOVA on forehand good serve</td>
<td>100</td>
</tr>
<tr>
<td>14. Mean comparison on forehand good serve</td>
<td>100</td>
</tr>
<tr>
<td>15. ANOVA on forehand scored serve</td>
<td>104</td>
</tr>
<tr>
<td>16. Mean comparison on forehand scored serve</td>
<td>104</td>
</tr>
<tr>
<td>17. ANOVA on forehand missed serve</td>
<td>105</td>
</tr>
<tr>
<td>18. Mean comparison on forehand missed serve</td>
<td>105</td>
</tr>
<tr>
<td>19. ANOVA on backhand good serve</td>
<td>106</td>
</tr>
<tr>
<td>20. Mean comparison on backhand good serve</td>
<td>106</td>
</tr>
<tr>
<td>21. ANOVA on backhand scored serve</td>
<td>107</td>
</tr>
<tr>
<td>22. Mean comparison on backhand scored serve</td>
<td>107</td>
</tr>
<tr>
<td>23. ANOVA on backhand missed serve</td>
<td>108</td>
</tr>
<tr>
<td>24. Mean comparison on backhand missed serve</td>
<td>108</td>
</tr>
<tr>
<td>25. ANOVA on return scored</td>
<td>109</td>
</tr>
<tr>
<td>26. Mean comparison on return scored</td>
<td>109</td>
</tr>
</tbody>
</table>
27. ANOVA on rallies scored .................................. 110
28. Mean comparison on rallies scored ....................... 110
29. ANOVA on returns missed ................................. 111
30. Mean comparison on returns missed ...................... 111
31. ANOVA on rallies missed ................................. 112
32. Mean comparison on rallies missed ...................... 112
33. ANOVA on total of returns ............................... 113
34. Mean comparison on total of returns .................... 113
35. ANOVA on total of rallies ............................... 114
36. Mean comparison on total of rallies ...................... 114
37. ANOVA on total of forehand serves ..................... 115
38. Mean comparison on total of forehand serves .......... 115
39. ANOVA on total of backhand serves .................... 116
40. Mean comparison on total of backhand serves .......... 116
41. ANOVA on total of serves ............................... 117
42. Mean comparison on total of serves ..................... 117
43. Major treatment effects on the variables in game play ........ 119
44. Mean comparisons of student perceptions on the three grading systems .................. 123
45. Principle component analysis of 15 questions ............. 125
46. Summary table on the effects of the three grading systems on different student outcome and process measures .................. 127
47. Mean comparisons of student perceptions on the three grading systems .................. 199
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Illustration of Design Model</td>
<td>63</td>
</tr>
<tr>
<td>2. Game Play Recording Sheet</td>
<td>76</td>
</tr>
<tr>
<td>3. Example of Game Play Recording and Computation</td>
<td>77</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Student achievement has been widely accepted as the primary goal of teaching and learning in any teaching and learning environment (Brophy, 1982; Brophy and Good, 1986; Dick and Carey, 1990; Gagne, Briggs, and Wager, 1988). Therefore, building a knowledge base about development of student achievement and enhancing practices related to student achievement become the two primary goals of research in education and teacher education (Brophy & Good, 1986; Siedentop, 1983a).

First, student achievement may be defined as learned capabilities such as intellectual skills, cognitive strategies, attitudes, verbal information, and motor skills, and include short and long term outcomes or products within either cognitive or affective domains (Dick & Carey, 1990; Gagne, et al, 1988). It should be noted that the above categories of student achievement are found to be weighted differently when related to different subject matter due to the nature of the different content areas. For example,
motor skill and fitness are unique factors in physical education and thus emphasized more. In contrast to physical education, a math teacher will focus more on the intellectual skill domain in terms of how to solve a math problem.

Second, in order to understand the mechanics of student achievement development, looking for and identifying the possible variables, behaviors, activities, and other relevant factors and trying to enhance student achievement become primary tasks for research in education in general and physical education in particular (Brophy & Good, 1986; Dunkin & Biddle, 1974; Siedentop, 1983a). It should be noted that the importance of the aforementioned variables, behaviors, activities, and other relevant factors, is imbedded in the effects and meanings they may have on student achievement or which may explain directly or indirectly the mechanics of student achievement development. In other words, given student achievement as the dependent variable, the major tasks for research become identifying independent variables which influence achievement and examining causal relationships between the two sets of variables (Brophy & Good, 1986).

Over the last two decades or longer, research conducted in general and physical education teacher education programs has focused on accomplishing these major tasks through

Although this task is far from complete, major strides (Locke, 1982) have taken place and include: process-product research (Rosenshine, 1979), student mediation studies (Berliner, 1979), classroom management and environment studies (Soar & Soar, 1979), ecological paradigm studies (Doyle, 1979), and those parts of studies and research related specifically with grading systems and policies (Boyce, 1990; Doyle, 1979; Harrison, 1983; Hensley, 1990; Son, 1989; Tousignant, 1982).

Recently, grading systems as part of teacher and student accountability systems have been addressed as an important controlling, mediating, and driving independent variable in terms of student learning (Boyce, 1990; Doyle, 1979; Harrison, 1983; Hensley, 1990; Son, 1989; Tousignant, 1982). Differences in grading systems in physical education at elementary, secondary, and college and university levels have been described with the meanings and importance of these grading systems discussed (Jones, 1989; Son, 1989; Tousignant, 1982). The relevant literature on grading could be briefly summed up as follows: Grading is the process of assigning marks (grades) to students based upon assessment of changes in student behavior or some other criteria.
(Boyce, 1990; Harrison, 1983; Hensley, 1990). Thus, the grading system refers to a set of processes (Harrison, 1983). Other related concepts are accountability system, formal accountability, and informal accountability (Doyle, 1979; Son, 1989). The importance of grading is widely stressed as related to program effectiveness and educational quality (Hensley, 1990; Miller, Dowell, and Pender, 1989; Summerfield, 1990). Evaluation is generally thought of as an integral component of the instructional process and is used not only for assigning grades, but for maintaining program effectiveness and for providing curriculum justification (Dunham, 1986; Safrit, 1990).

Assigning grades has motivating, informative, and diagnostic functions (Safrit, 1973). Motivating function means that grading stimulates, directs, and guides student learning efforts (Summerfield, 1990). Informative function provides information on the progress of students toward program objectives to students, teachers, and parents as well as administrators (Harrison, 1983). Finally, the diagnostic function informs teachers, students, parents, and administrators about the current condition of student learning and identifies problem areas (Harrison, 1983).

Among the three grading functions, motivation is the theoretical underpinning for this study. Grading as a motivating and driving force as mentioned above has been
well documented in the literature (Boyce, 1990; Doyle, 1979; Harrison, 1983; Loucks, 1987; Son, 1989; Tousignant, 1982). Some assumptions about grading as a motivating and driving force have been stressed by motivation theory (McGregor, 1960) and the writings of Doyle (1979). Motivation Theory X suggests that due to the nature of students they do not like studying and learning unless forced to do so (Hersey & Blanchard, 1972; Sergiovanni, Burlingame, Coombs, & Thurston, 1987). In Doyle's work, grading is considered an essential component related to effective teaching and learning. Grading has the power to guide, direct, motivate, and drive student learning toward desired student learning in specific ways. According to Doyle (1979), students initially are not likely to be interested in the learning process, instead, students are pursuing the results of learning. The result of the learning activity of students in schools or universities is represented by grades. The major task of students therefore is defined as an exchange of performance for grades. That students care more what grades they receive than actual achievement they obtained indicates the power and likelihood of using grading as a driving force. Grades are a radical indicator of current student status and also influence a student's future in terms of school and employment opportunity.
Second, Doyle (1979) states that the school environment is problematic from an individual student point of view due to the characteristics of current schools: the range of purposes, number of participants, appropriateness of curricula, forms of instruction, quality of teaching, and duration of existence. Therefore, students are likely to dislike what teachers intend to teach unless they are forced to do so (Doyle, 1979). More important, in order to secure a successful exchange of performance for grades, students have developed individual competencies by strategically and skillfully following directions guided by grading systems and by meeting the criteria set by the grading systems.

This notion has also been well documented in research on teaching physical education (Boyce, 1990; Harrison, 1983; Son, 1989; Tinning, 1983; Tousignant, 1982;). Tousignant (1982) found that when there was no grading of students' performance, the instructional task system was suspended. She also found that the actual tasks and the functional curricula were greatly affected by the grading system.

It is widely accepted that grading components should follow the objectives of instruction (Boyce, 1990; Dick & Carey, 1990; Gagne, Briggs, & Wager, 1988; Harrison, 1983). However, it appears that functionally, the grading components become the actual objectives of instruction. In other words, the content the grading system focuses on
becomes the functional curricula. What students learn is not what the teacher taught but what the grading system assesses (Doyle, 1979; Tousignant, 1982). Therefore, grading components merit special attention. There are many grading components in physical education which are reported although there is not total agreement (Boyce, 1990; Fox, 1959; Harrison, 1983; Hensley, 1990; Larson & Yocom, 1951; Lumpkin & Avery, 1986; Pangrazi & Darst, 1985; Tousignant, 1982).

Theoretically, physical education affects all capability domains; motor skill, intellectual skills, cognitive strategies, verbal information, and attitudes (Gagne, et al., 1988; Siedentop, 1983a). The notion that physical education impacts all human capability domains is well illustrated in the grading components.

Some researchers (Fox, 1959; Hensley, East, Aten, Lambert, Baumgartner, & Stillwell, 1989; Son, 1989; Tousignant, 1982) indicate that few tests have been given to evaluate the degree to which student performance changes. The reality is that many teachers use subjective judgment of effort, attendance, sportsmanship, and performance to assign grades and to ascertain the effectiveness of an instructional program. However, it should be noted that the major characteristic of grading components in physical education is its variety (Hensley, 1990; Hensley, Morrow, &
In conclusion, grading systems have been considered an essential and powerful variable in driving, mediating, and directing student learning. Great variety is found in grading systems and grading components among teachers. It seems that there is a lack of knowledge about the nature of the grading system as a driving and motivating force. Furthermore, almost all findings on grading systems have been descriptive. Knowledge about the effects of grading systems on student learning is still limited primarily to some qualitative descriptions. Only two experimental studies on the effects and the mechanics of grading systems in terms of the roles of independent variables on student achievement have been reported (Boyce, 1990; Mueller, Aicinena, Corso, & Phillips, 1988). Also, research questions related to grading systems, which could be most appropriately answered by experimental methodology, have been raised. For example, what do different grading systems mean to students given other process variables unchanged (Boyce, 1990; Tousignant, 1982)? In other words, it should be meaningful to test the effects of different grading systems on student achievement such as skill improvement. What is the nature of a grading system in terms of student achievement? Can we answer these questions through
experimental research? For example, what are the effects of a managerial grading system, in which attendance and appropriate participation are the foundation for grading, on student outcomes which are related to motor skills, intellectual skills, verbal information, and attitudes?

**Statement of the Problem**

The purpose of this study was to identify and test the effects of three different grading systems on learning outcomes and other factors of students in three table tennis classes in a university Basic Instruction Program. The specific purposes of this study were stated as the following research questions:

1. What were the effects of three different grading systems on student behaviors measured by a systematic observation instrument?

2. What were the effects of three different grading systems on student attendance rates?

3. What were the effects of three different grading systems on the scores of two skill tests?

4. What were the effects of three different grading systems on the scores of a written examination?

5. What were the effects of three different grading systems on the scores of game play variables?

6. What were the effects of three different grading systems on the scores of an attitude questionnaire.
Null Hypotheses

Six null hypotheses were tested that there would be no differences on learner outcome and process measures according to grading treatment condition; Skill, Managerial, and Typical. More specifically, the grading system would not influence a difference:

1. on student behaviors as measured by the systematic observation instrument.
2. on student attendance rates.
3. on the scores of the two skill tests.
4. on the scores of the written examination.
5. on the measures of game play variables.
6. on the scores of an attitude questionnaire.

Definitions

Grading is not synonymous with the interpretation of evaluation, it is narrowly considered as the process of assigning marks (grades) to students based upon assessment of changes in student behavior or other criteria (Boyce, 1990; Harrison, 1983; Hensley, 1990).

Grading Components refer to those student learned capabilities such as motor skills, intellectual skills, cognitive strategies, attitudes, and verbal information on which student grades are based (Dick and Carey, 1990; Gagne, et al., 1988).
Grading System refers to a set of processes used in assessing performance according to selected criteria (Harrison, 1983).

Formal Accountability refers to a group of consequences affecting grade-exchange contingencies specified and communicated in advance of the performance (Son, 1989).

Accountability Systems refer to sets of practices teachers use to establish and maintain student responsibility for work (Worsham and Everstion, 1980).

Managerial Grading System refers to that grading practice in which student grades are assigned primarily based on student activity, appropriate attendance, and participation such as being physically present, dressing appropriately, being punctual, remaining until class is over and staying on task.

Skill Grading System refers to that grading practice in which student grades are given based on student achievement on two skill tests; one-minute volley test and a serve test, and the scores on skills demonstrated in game play.

Typical Grading System refers to that grading practice in which student grades are assigned; 40% on skill tests, 40% attendance and participation, and 20% on a written examination.
Delimitations

1. Subjects consisted of students enrolled in three Table Tennis I classes during Spring, 1991 at The Ohio State University.

2. The three table tennis classes were randomly assigned to one of the three treatments.

3. Duration of treatment was ten weeks covering the entire Spring Quarter, 1991.

4. The three intact classes were taught by the same teacher who was the investigator.

5. Equipment used in these three classes was provided by the university although students were allowed to use their own paddles and balls.

6. The three classes were taught in an indoor multipurpose room.

7. Only one camera with a wide angle lens was used to conduct permanent recording for each class.

8. Every two students had at least one table. All of the tables were arranged in a line with about two-three yards between tables.

9. Illumination in the room was considered normal although some areas may have been lighter.

10. Student skill achievement was tested by two skill tests.
11. Student cognitive achievement was measured by two written tests.

13. Student attitudes toward the three grading systems were measured by a 15-item questionnaire.

Limitations

1. Assignment of the three intact classes to each treatment may have caused some threat to the internal validity of this study.

2. The researcher, as the teacher for the three classes, may have had some effect due to reactivity which the teacher behavior congruence instrument (DIBA) may have failed to identify.

3. Duration of treatment may not have been appropriate.

4. Use of one camera to produce permanent recording products may have missed some information which may have been caught using more than one camera from more than one angle.

5. Different tables, balls, and paddles may have had unidentified effects on the dependent variables.

6. Overall student skill may not have been well represented by the measures on the two skills tests.

7. Student cognitive achievement may not have been well represented by the scores on the written examinations.
8. Student general attitude toward the three grading systems may not have been validly measured by the 15-item questionnaire.

Significance of this Study

Student achievement has been widely accepted as the primary goal of teaching and learning in any teaching and learning environment (Brophy, 1982; Brophy and Good, 1986; Dick and Carey, 1990; Gagne, Briggs, and Wager, 1988). Given student achievement as the dependent variable, knowledge on the nature and functions of those robust independent variables which have strong effects on student achievement is obviously essential to enhance student achievement in physical education.

Recently, grading systems as part of teacher/student accountability have been addressed as an important controlling, mediating, and driving independent variable in terms of student achievement (Boyce, 1990; Doyle, 1979; Harrison, 1983; Hensley, 1990; Son, 1989; Tousignant, 1982). It has been well documented that grading systems and grading components in practice vary tremendously (Boyce, 1990; Fox, 1959; Hensley, et al., 1989; Mathews, 1978; Miller, et al., 1989). Although researchers (Brophy, 1982; Brophy and Good, 1986; Dick and Carey, 1990; Gagne, Briggs, and Wager, 1988) believe that grading has important guiding, driving, and motivating functions in terms of student achievement, very
few functional relationships between grading systems and student achievement have been identified. In other words, although grading has been a basic component in school practice for a long time, people have only recently started to address its motivating function. Very little knowledge about the nature of grading systems based on research is available.

The objective of this study was to identify the effects of three common grading systems, Skill, Managerial, and Typical on student motor skills, intellectual skills, cognitive strategies, verbal information, and attitude. It is expected that results from this study will shed light on the nature of a grading system as a driving, guiding, and motivating force and will add new information to the knowledge base of accountability in terms of student achievement in physical education. More specifically, it is expected that specific functional relationships between the three grading systems and student skill, cognitive, and process measure achievements as specified by this study will be established.
CHAPTER II
THE REVIEW OF RELATED LITERATURE

A review of the literature related to grading in education and physical education is presented. This chapter is organized into three sections: important assumptions and theories of research in teaching as it relates to grading, related research paradigms in education and teaching in physical education, and research on grading in physical education.

Important Assumptions and Theories Related to Grading

The areas covered in this section are related to the notion of student achievement or student learned capabilities as the dependent variable in research on teaching and on teaching physical education, motivation theory X, and Dunkin and Biddle's (1974) Model.

Student Achievement as A Dependent Variable

Identifying dependent variables and the assumptions surrounding them are primary concerns in teacher effectiveness research in both general and physical education (Brophy & Good, 1986; Siedentop, 1983a). Student
achievement has been used as a primary dependent variable in teacher effectiveness research (Brophy, 1982; Brophy & Good, 1986) and research on teaching in physical education (Locke, 1982; 1986; Nixon & Locke, 1973; Placek & Locke, 1986; Siedentop, 1983a; 1983b). Student achievement may be related to cognitive and affective aspects or they may be categorized within five domains: intellectual skills, cognitive strategies, verbal information, attitudes, and motor skills according to Gagne, Briggs, and Wager (1988). They refer to such student achievement in these five domains as learned capabilities.

Taking student achievement as the dependent variable in teacher effectiveness research and research on teaching in physical education implies the use of student achievement as a criterion of effective teaching in both classrooms and gymnasiums (Brophy & Good, 1986; Locke, 1986; Siedentop, 1983a). It has been shown that physical education has an influence on student learning capabilities in all five domains (Mao & Zakrajsek, 1988). Mao and Zakrajsek (1988) surveyed 475 secondary physical education teachers in Idaho, Montana, Oregon, and Washington using a 14-item opinionnaire which was developed after careful review of the goals and objectives of physical education expressed in the literature. The 14 goals and objectives accepted and ranked by those teachers in order were: fitness and wellness, fun
and enjoyment, attitude for active adult lifestyle, fair play/sportsmanship/character building, learning motor skills, social values/team relationship, recreation for leisure time, success in skill development, cooperation/competition, learning many games/activities, health/diet, moral development/ethical behavior, knowledge of game rules/tactics/insights, and principles of movement/body relationships. It is not difficult to see that these goals and objectives fall into the five domains. Therefore, samples of student achievement in all possible domains in a research study is desirable (Baumgartner & Jackson, 1982; Bayless, 1978). However, it should be pointed out that there are two kinds of measures in terms of student achievement: process measures and product outcome measures (Alexander, 1983; Anderson, 1983; Berliner, 1979; Gagne, et al. 1988; Griffey, 1983; Locke, 1986; Pangrazi & Darst, 1985; Pieron 1982; Rife, Shute, & Dodds, 1985; Siedentop, Tousignant, & Parker, 1983; Wuset, Mancini, Mars, & Terrillion, 1986). Because both process and product measures will be employed in this study, the advantages and disadvantages related to these two kinds of measures in the literature are briefly presented.

Process measures refer to using proxy or criterion variables such as time on task (TOT) or academic learning time (ALT) as the measures since TOT or ALT are so defined
that their nature is close to student achievement (especially, for the cognitive domain). In addition, they are consistently found to be positively related with student learning (Berliner, 1979; Siedentop, Tousignant, & Parker, 1982). One of the advantages of using process measures is that they are content valid in the sense that TOT or ALT are always related with specific subject matter and TOT/ALT measures learning during the on-going process (Siedentop, et al., 1982). Process variables also provide ongoing information about the teacher's teaching and students' learning which negates waiting until the end of a unit or a course when product test scores become available (Berliner, 1979). However, TOT/ALT are always limited by their nature of temporal recording. For example, TOT/ALT does not distinguish whether a student is on-task or off-task when the student appears to be attending to instruction, lecture, or demonstration. In other words, there is the same difficulty with TOT/ALT in measuring the "competent bystander" in both the classroom or gymnasium (Anderson, 1983; Siedentop, et al., 1982).

Concern should be given to subject-specific matter when using process measurements (Anderson, 1983). Other variables used as process measures are opportunity to respond (OTR) and content covered which may be better process measures but usually more inconvenient or difficult
to conduct (Anderson, 1983; Brophy, 1986). Finally, process measures primarily focus on the cognitive domain in classrooms and motor domain in physical education, therefore they do not differentiate student gains according to different domains (Griffey, 1983).

Product measures such as standardized achievement tests for math or reading have been used as another kind of teacher effectiveness outcome measure (Brophy & Good, 1986; Gagne, et al., 1988). The advantage is that they aim at measuring student achievement directly in specific areas, (e.g., fourth grade reading, acquisition of a novel golf skill, or a certain kind of attitude). Results of these tests may be interpreted in setting a normal-curve. However, they have their weaknesses in several ways. First, there is a content-valid concern (Siedentop, et al., 1982). The product measure test, especially a standardized test, may not measure student outcomes in a way that is consistent with the content taught by the teacher or in the methods used by the teacher. This is well illustrated by Berliner's (1979) Beginning Teacher Evaluation Study (BTES). For example, if teachers let their students spend different amounts of allocated time, engaged time, and time on task (TOT) for different subject areas, it is difficult to judge these teachers' effectiveness by comparing results on a standardized test. In addition, it is important to sample
all aspects of content taught in a product test, but it is not easy to do (Terry & Safrit, 1990). For example, most product measure tests are related to the cognitive domain and short-term effects (Brophy, 1982; Brophy & Good, 1986; Rosenshine, 1979; Siedentop, 1983a). It should be noted that product measurement tests in the affective domain are limited primarily to student self reports methods (Siedentop, 1983a).

In short, student achievement has been taken as the dependent variable. Student achievement can be categorized into five domains; intellectual skill, cognitive strategies, verbal information, motor skill, and attitude. Although both process measures and product measures have advantages and disadvantages, sampling student achievement in all possible domains is desirable in teacher effectiveness research and research in teaching physical education.

Motivation Theory X

Assumptions about the nature of man plays an important role in understanding student learning (Hersey & Blanchard, 1972; Sergiovanni, Burlingame, Coombs, & Thurston, 1987). Recently, the position in research on teaching in physical education that assumes teaching and learning are natural phenomena has been widely challenged (Locke, 1989; Popkewitz, 1984; Schempp, 1987; 1988). It is likely that students possess relatively stable internal conditions such
as attitudes, concerns, needs, motivations, and other learned capabilities, which impact on their behaviors in both the classroom and gymnasium (Dick & Carey, 1990; Gagne, Briggs, & Wager, 1988; Locke, 1989; Schempp, 1987; 1988). These internal conditions have been well explained by motivation theorists. There are however many motivation theories that hold different positions (Hersey & Blanchard, 1972; and Sergiovanni, Burlingame, Coombs, & Thurston, 1987). One of these motivation theories, Theory X (McGregor, 1960), appears to be most closely related theoretically to this study and will be briefly reviewed.

Theory X provides a list of assumptions about the nature of man in society. These assumptions according to Hersey & Blanchard (1972) and Sergiovanni, et al., (1987), may be interpreted in the instructional setting as follows: (a) The average student has an inherent dislike of work and will avoid it; (b) because of a dislike of study, most students must be coerced, controlled, directed, or threatened with punishment to get them to put forth adequate effort toward the achievement of instructional objectives; (c) the average student prefers to be directed, wishes to avoid responsibility, has relatively little ambition, and wants security above all.

In short, it is not difficult seeing that Theory X partially explains why the use of grades becomes the
instructional weapon for prodding students toward higher levels of achievement.

**Dunkin and Biddle Model**

Given student achievement as the dependent variable in research and the goal of instruction, a useful model for enhancing student achievement and understanding this dependent variable is provided by Dunkin & Biddle (1974). This model suggests four kinds of variables: presage variables, context variables, process variables, and product variables.Presage variables refer to teacher formative experiences (social class, age, sex...), teacher training experiences (university attended, training-program features, practice-teaching experiences...), and teacher properties (teaching skills, intelligence, motivations, personality traits...). Context variables refer to student variables (social class, age, sex, abilities, knowledge, attitudes...), school and community contexts (climate, ethnic composition of community, bussing, school size...), and classroom contexts (class size, textbooks, educational television...). Process variables contain both teacher and student classroom behaviors and their interactions. Finally, product variables are classified as immediate student growth (subject-matter learning, attitudes toward subject, growth of other skills...) and long-term student effects (adult personality, professional or occupational skills...).
According to this model, product variables are produced by the process variables and process variables are impacted heavily by presage and context variables.

It should be noted that this model provides a framework for doing research using process-product method although this model is unable to include a grading system that explicitly acts as the independent variable because of the influence of student mediation.

In short, Dunkin and Biddle (1974) provide a conceptual framework in which student achievement is taken as a dependent variable which is related to other independent variables of which grading could be added.

Research Paradigms Related to Grading

These paradigms contain process-product, ecological, and experimental teaching unit research studies in physical education.

Process-Product Paradigm

The review on process-product research is organized as: the major research done, the major findings which are primarily consistent in research studies, the variables studied but with inconsistent or non-general results, and the major characteristics of methodology. The purpose of this part of the review is to show the relevance of grading as a possible independent variable, which, has not been studied yet in the process-product paradigm. In addition,
this part of the review intends to show how grading fits as an independent variable within the process-product paradigm.

It is safe to say that process-product research represents the most often used paradigm in research on teaching in terms of the quantity of findings and number of research studies conducted (Brophy & Good, 1986; Wittrock, 1986). Also, process-product research has a strong impact on other research paradigms by both providing research findings and stimulating research questions (Brophy & Good, 1986; Wittrock, 1986).

The major research question pursued in the process-product research paradigm can be viewed at two levels. First, in a narrow perspective, process-product research studies the linkages between variables of teacher behaviors and variables of student achievement. Through a wider point of view, process-product research looks at possible linkages between those variables related to teaching and those variables related to learning (Brophy & Good, 1986). Literature in process-product research studies have been well documented (Berliner, 1976; Brophy, 1982; Centra and Potter, 1980; Doyle, 1977; Gage, 1978; Good, 1979; Rosenshine & Stevens, 1984).

According to the inclusion criteria of Brophy and Good (1986), the major process-product research studies have been primarily reported by Brophy and Evertson (1974), Flanders
(1970), Fortune (1967), Good and Grouws (1975), McDonald & Elias (1976), Rosenshine (1971), Soar and Soar (1968), Stallings and Kaskowitz (1974), and Wright and Nuthall (1970). The major findings which were found to be consistent in earlier process-product studies were initially summed up by Rosenshine and Furst (1971) in their famous ten-item list:

- clarity of the teacher's presentation,
- enthusiasm of the teacher,
- variety of activities during the lesson,
- task-oriented and businesslike behaviors in the classroom,
- content covered by the class,
- the teacher's acknowledgement and encouragement of students' ideas during discussion,
- criticism of the student (negatively related to achievement),
- use of structuring comments at the start of and during a lesson,
- use of a variety of types of questions,
- probing of students' responses by the teacher.

However, fifteen years after the Rosenshine and Furst reporting, Brophy and Good (1986) provided another list of consistent findings derived from process-product research. Their list suggests that an effective teacher is positively related to:

- more opportunity for students to learn and more content covered,
- high expectations from teacher and high amount of allocated time,
- good classroom management and high student engaged time,
- consistent student success and high academic learning time,
- active teaching,
- structured presentation,
- redundancy and sequencing of important points,
- clarity of presentation,
(i) enthusiasm of teacher, (j) clarity of question, (k) appropriate postquestion wait-time, (l) appropriate selection of the respondent, (m) waiting time for the student to respond, (n) acknowledgement of correct responses, (o) appropriate reaction to partly correct responses and incorrect response, (p) proper reaction to student questions and comments, and (q) appropriate handling of seatwork and homework assignments.

According to Brophy and Good (1986), there are other important variables studied in process-product research although no consistent results have been reported. These important variables include: size of class, difficulty level of questions, cognitive level of questions, grade level, student social economical statues, student ability, student affect, and teacher's intentions and objectives.

Within the process-product paradigm, the progressive loop of descriptive-correlational-experimental research has been completed (Brophy and Good, 1986). This loop means that people start by describing what is going on, establishing the association between possible teaching variables and student achievement variables, and finally, based on these previous two variables, experimental studies are conducted to test specific functional relationships between or among sets of teaching variables and student achievement variables. These relationships are usually the
research questions raised in the correlational or descriptive studies. This loop implies a methodological and progressional development in the process-product research paradigm.

However, it should be noted that the methodology employed in process-product research is related to the natural science methodology which has its limitations in studying teaching (Schempp, 1987). For example, it is widely believed that learning occurs on an individual basis whereas the units of analysis in process-product research are usually dealing with measures of an entire class (Brophy and Good, 1986; Schempp, 1987). In addition, it is believed that process-product research ignores those variables related to student mediating activities (Doyle, 1979).

In short, the process-product research paradigm provides a fruitful model for studying teaching. According to this model, those variables related with teaching are taken as independent variables while those variables related with student achievement are considered dependent variables. Many independent variables have been employed, however, employing a grading system as the independent variable has not been reported in the process-product literature.

Ecological Paradigm

The ecological paradigm can be viewed as a post-process-product research paradigm because of the realization
of the major limitations in process-product research (Doyle, 1979; Jones, 1989). Ecological researchers believe that the reality of the classroom and gymnasium is far more complicated than a set of linkages between two sets of variables as treated by the process-product paradigm (Doyle, 1979; Jones, 1989). According to Doyle (1977, 1986), the complexities of a classroom environment could be described by six components, which are cited by Jones (1989):

- Multidimensionality which simply means that a large quantity of events and tasks take place in classrooms.
- Simultaneity refers to the occurrence of many things at the same time in classrooms.
- Immediacy describes the rapid pace of events which occur in classrooms.
- Unpredictability refers to the very nature of classrooms which dictates that events often take unexpected turns.
- Publicness indicates that classroom events occur in front of many people.
- History implies that throughout the school year, classrooms accumulate a common set of experiences and routines which provide a foundation for conducting activities (p. 26).

Therefore, a framework for understanding classrooms needs to include at least the following four features (Doyle, 1979): first, a reciprocal causality between students and teachers in the classroom environment should replace the unidirectional assumption related to the process-product paradigm, second, an information-processing view of the mediational strategies students use to navigate classroom environments should be built into this new framework, third, in addition to taking an entire class as the unit of analysis, more attention should be given to
individual students, and finally, a systemic view of the natural classroom environment should be used. In other words, this new framework has three basic dimensions: (a) a naturalistic dimension which implies the goal of ecological analysis that allows the reality of the classroom to impinge upon the investigator's subjectivity until the categories for description are determined by the phenomena of the classroom itself, (b) an emphasis on environment-behavior relationships, and (c) a focus of ecological analysis on the functional value or adaptive significance of behaviors in an environment.

Furthermore, from a student's point of view, Doyle (1979) believes that the complexities of the classroom environment could be viewed through three aspects:

First, there appears to be an abundance of information resources which are usually beyond the control of a student. Second, "not all elements of the classroom are equally reliable as instructional cues" .... For example, teacher instructions are sometimes ambiguous and incomplete. Finally, "since the classroom is a mass-processing system, the degree to which the level of information and flow of activities necessarily match the individual student's interests or abilities is limited" (p. 193).

In order to understand and analyze the complexities of classrooms, notions of tasks and task structures were created (Doyle, 1979; 1986; Jones, 1989). A task refers to what a person must do to successfully meet the demands of the situation (Jones, 1989). Task structures refer to
patterns for organizing and implementing subject matter and non-subject matter activities. However, "with special reference to academic outcomes, the formal structure of classroom tasks can be defined as an exchange of performance for grades" (p. 192). Each pattern consists of a goal and operations to achieve that goal (Doyle, 1979). The importance of these two notions is further explained by Doyle (1979):

The problems of learning from classrooms would seem nearly insurmountable if one were to assume that the environment is composed of discrete entities. There would seem, however, to be a system of overlapping task structures that integrates elements of the classroom. Each of these task structures consists of a goal and a set of operations to achieve that goal, ..., each task structure defines a behavior ecology (p. 192).

In order to fulfill their task in terms of exchange of performance for grades, students develop their classroom competencies to deal with the complexities of the classroom environment by decreasing the degree of ambiguity and risk in accomplishing tasks (Doyle, 1979; Tousignant, 1982). Competency in using and dealing with accountability that occurs when one person is keeping an account of the behavior of another to see whether it meets specifications (Son, 1989), becomes particularly crucial to both teachers and students (Doyle, 1979; 1986; Tousignant, 1982). In order to complete their task of exchange of performance for grades, students are likely to only concentrate on those tasks which
relate to their grades (Doyle, 1979).

In summary, the ecological paradigm could be viewed as a post-process-product conceptual framework which emphasizes the complexities of the classroom environment. Task structures are an important notion in understanding the complexities of the classroom environment. Furthermore, accountability has crucial meaning for accomplishing tasks in classrooms.

Experimental Teaching Units (ETU) in Research on Teaching in Physical Education

This part of the review consists of backgrounds, concepts, and assumptions behind ETU studies, the major findings, characteristics of ETU methodology, and its limitations. The purpose of this part of the review is to show what has been done in research on teaching in physical education by using ETU to help readers see the relevance and meaningfulness of an experimental design study using grading as an independent variable in research on teaching in physical education.

ETU studies are usually based on the information, knowledge, experience, and questions raised from other studies such as descriptive and correlational studies. In teacher effectiveness research, experimental studies are considered as the last stage in the descriptive-correlational-experiment loop (Brophy & Good, 1986).
An Experimental Teaching Unit (ETU) refers to a standardized unit of instructional content presented to students in a short instructional period (Metzler, 1983). The primary purpose of an ETU study is to test the relationship between a few independent variables and a few dependent variables or between one independent variable or a few independent variables and one dependent variable in a well-controlled experimental environment. The independent variable is assigned to the ETU and the effects of the independent variables on the dependent variables are measured. Results of the measurement provide information about the relationship between the independent and dependent variables (Paese, 1986).

The major findings of ETU studies in research on teaching in physical education could be viewed from four areas: student utilization of time, instruction variables, teacher feedback, and teaching approaches.

First, student utilization of time may be the area where ETU studies have the most consistent findings (Locke, 1986; Paese, 1986). Motor-ALT was found positively related with student motor skill achievement (Metzler, 1983) and the students of more effective teachers have larger amounts of time on task than those students taught by less effective teachers (Graham, Soares, & Harrington, 1983). However, in the above studies, student engaged time was not different in
student motor skill achievement between more effective teachers and less effective teachers. It should be noted that the above findings are parallel with those found in general teacher effectiveness studies (Brophy and Good, 1986). Second, instructional variables in ETU studies accounted for little student achievement (Pieron and Graham, 1984). Explanations for the small contribution of instructional variables to student achievement were addressed by Paese (1986) as follows: 1) the nature of physical education is body movement, where physical demonstration may be the more economical and better communicating means, 2) the instruction time in those ETU studies is too short to produce measurable effects, and 3) the quality of the teacher presentation and appropriateness of the instruction may be problematic.

Third, there were inconsistent findings related to feedback in ETU studies (Graham, et al, 1983; Paese, 1986). On one hand, detailed informative feedback was found negatively related to student achievement while feedback to total movement was positively related to student achievement (Paese, 1986). It is interesting that more effective teachers were found to provide slightly less feedback. Affective feedback was the most frequently used type of feedback (Paese, 1986).
According to Paese (1986), possible explanations about their findings are: 1) no single variable was powerful enough to discriminate more/less effective teachers; 2) length of treatment time might not provide the specific points in the learning process for students to receive feedback; 3) tasks used in those ETU studies were simple. These tasks were relatively simple and novel, therefore, knowledge of both results and performance were readily available, and 4) it was found that there were no differences in student skill acquisition and a student rating of self-efficacy when using different teaching approaches (Salter & Graham, 1985).

Finally, one conclusion from the major findings in ETU studies was that there were no significant effects of independent variables on the dependent variables (Graham, et al, 1983; Metzler, 1983; Paese, 1986).

A brief summary of the characteristics of ETU methodology are: 1) the information of ETU studies is based on the measurement of a pre- and posttests, 2) the sample size is small and the time of treatment in effect is short, and 3) only one or a few of those variables and factors which may have contributed to the measurable dependent variables are included. However, it should be noted that the above characteristics of ETU studies embody both its advantages and disadvantages because studies like those
experimental studies completed in general teacher
effectiveness research are obviously expensive to conduct in
terms of cost, effort, and time (Graham, et al, 1983;

In summary, ETU studies represent the final step of the
descriptive-correlational-experimental loop on research in
teaching physical education. Both the advantages and
disadvantages of an ETU study exist in its short duration of
treatment, low cost, and small scale nature. So far, the
contributions of ETU studies are limited within replications
of some results done in the other research paradigms.

Research on Grading in Physical Education

This section will present research in teaching physical
education which is directly related to the notion of
grading. This section is organized as follows: grading of
physical education in measurement/evaluation and in sport
pedagogy, concept of a grading system, importance of
grading, functions and purposes of grading, grading system
as driving force, grading foci, gap between theory and
practice in grading, difficulties in implementing grading
in the real world, and other issues related to grading in
physical education.
Grading of Physical Education in Measurement/Evaluation and Sport Pedagogy

The issue of grading can be seen through two perspectives in terms of (a) grading in measurement and evaluation in physical education and (b) grading in sport pedagogy. Traditionally, grading is a major topic in the subject area of measurement and evaluation in physical education (Barrow & McGee, 1971; Clarke & Clarke, 1987; Kirkendall, Gruber, & Johnson, 1987; Larson & Yocom, 1951; Pangrazi & Darst, 1985; Mathews, 1978; Safrit, 1973; Singer, 1967; Solley, 1967). Topics in this area are broad from test construction to program evaluation, from the practice of grading to theory of measurement, from goals and measurable objectives to statistical techniques.

Recently, grading has obtained attention in sport pedagogy in the sense that the issue of grading is viewed primarily in light of student behavior and learning (Boyce, 1990; Lumpkin & Avery, 1986; Miller, et al, 1989; Mueller, Aicinena, Corso, & Phillips, 1988; Son, 1989; Tousignant, 1982). There are two important themes in research on grading in this area. The first is related to the notion of "exchange of performance for grade" (Son, 1989; Tousignant, 1982). The next one is related to the conceptual framework of the process-product paradigm because this theme considers grading as the independent variable and other student
outcomes as dependent variables (Boyce, 1990; Mueller, et al., 1988). For example, Mueller, et al. (1988) believe that grading is one of the most important independent variables for skill improvement.

Numerous factors have been found to affect skill improvement. Some of the factors that have been reported are: amount of practice, type of practice, amount and type of feedback, knowledge of the task requirements, intent to improve, and stress (p. 30).

The Concept of Grading

Grading is the process of assigning marks (grades) to students based upon the assessment of changes in student behavior or some other criteria (Boyce, 1990; Harrison, 1983; Hensley, 1990) while grades refer to evaluations of a student's performance which become a part of the permanent record (MacFarlane, 1981). Thus, the grading system refers to a set of processes in grading (Harrison, 1983). Other related concepts here are accountability system, formal accountability, informal accountability (Son, 1989), and evaluation. Formal accountability refers to a group of consequences affecting grade-exchange contingencies specified and communicated in advance of the performance (Son, 1989). Evaluation is the process of obtaining information (data) and using it to form judgments which in turn are used in decision making (Dunham, 1986). Evaluation is also the process of collecting data from students which can serve as the basis for evaluating the effectiveness of
instruction as well as for assigning grades (Imwold, et. al., 1982).

The Importance of Grading

The importance of grading is widely stressed as related to program effectiveness, appropriateness of the curriculum, the magnitude of student achievement, and educational quality (Hensley, 1990; Miller, Dowell, & Pender, 1989; Dunham, 1986; & Summerfield, 1990). Evaluation is generally thought of as an integral component of the instructional process and is used not only for assigning grades, but for maintaining program effectiveness and for providing curriculum justification (Dunham, 1986; Safrit, 1973; 1990). Therefore, grading is one of the most difficult, yet most important, tasks with which teachers are presently confronted and each teacher should develop a sound and valid grading system (Dunham, 1986).

A student's grades are important because they determine to some extent whether this student will secure a good position, be admitted into college, win a scholarship, or gain prestige among his/her classmates (Neilson & Jensen, 1972). As a result of the importance of grading in physical education, a large part of the focus in physical education during the last two decades has been on evaluation (Bayless, 1978; Hensley, et al., 1989). The importance of evaluation can also be seen in that almost all students in teacher
preparation programs are required to take a course in testing, measurement, and/or evaluation. In addition, almost every textbook for courses on teaching, curriculum, and methods contains a section on evaluation (Dunham, 1986).

Furthermore, the importance of grading in physical education is related to accountability because grading is a radical part of an accountability system in physical education (Hensley, et al., 1987; 1989; Miller, et al., 1989). While the role of physical education in the public school curriculum has been questioned and challenged, ironically, physical educators typically are held responsible for the lack of fitness among the youth of the nation (Hensley, et al., 1989). The solution to the above dilemma is the evidence of effectiveness in physical education. In other words, an effective accountability system is needed to show that the instructional goals and the needs of students in physical education are accomplished and met. Obviously, grading is related to the radical parts of accountability in physical education (Boyce, 1990; Hensley, et al., 1987; 1989; Miller, et al., 1989).

Hensley, et al. (1987), thus, believe that it is likely that the survival of physical education may, to some extent, depend upon the efficacy of measurement and evaluation efforts. The significance of grading in physical education
could be further illuminated by revealing its functions and purposes.

Functions and Purposes of Grading

Assigning grades has motivating, informative, and diagnostic functions (Dunham, 1986; Neilson & Jensen, 1972; Safrit, 1973). Motivating function means that grading stimulates, directs, and guides student learning efforts (Summerfield, 1990). Informative function shows the progress of students toward program objectives to students, teachers, and parents as well as administrators (Harrison, 1983). For example, a grade tells a student how well his/her performance is. Finally, diagnostic function informs teachers, students, parents, and administrators in aspects such as what is the current condition of student learning, and what and where the problems are (Harrison, 1983). Among the three grading functions, the motivating one is the theoretical underpinning of this study, therefore the motivating function deserves careful reviewing.

Grading as Motivating and Driving Force

One of the functions of grading as a motivating and driving force for student learning process has been documented in the literature on physical education (Boyce, 1990; Harrison, 1983, 1987; Loucks, 1987; Mueller, et al. 1988; Neilson, 1972; Safrit, 1973; Tousignant, 1982). Some assumptions undergirding the notion of grading as a
motivating and driving force have been originally stressed by motivation theory and Doyle as reviewed in previous sections. Safrit (1973) believes that grading has the power for guiding, directing, motivating, and driving student learning toward the desired student and learning manner, direction, and specific ways.

Some further explanations about grading as a driving force have been provided by Son, 1989; Tinning, 1983; Tousignant (1982). First, students are not likely to be interested in the learning process, instead, students are pursuing the results of learning. The result of the learning activity of students in schools or universities is grades. The major task of students is defined as an exchange of performance for grades. That students care more about what grades they get than the actual achievement they obtain indicates the power and likelihood of using grades as a driving force. Second, the school environment in physical education is problematic from an individual student point of view due to the characteristics of current schools: the range of purposes, number of participants, appropriateness of curricula, forms of instruction, quality of teaching, and duration of existence, especially, the different demands of different tasks in physical education. More important, in order to secure a successful exchange of performance for grades, students have developed their individual
competencies by strategically and skillfully following the direction guided by grading systems and by meeting the criteria set by grading systems.

Tousignant (1982) found that when there was no grading on the students' performance, the instructional task system was suspended. She also found that the actual tasks and the functional curricula are greatly affected by a grading system other than what the instruction is like.

Boyce (1990) conducted an experimental study to assess the impact of grading practices (grading foci) on student skill performance of nine riflery classes at a university by employing three different grading systems as independent variables and student shooting scores as the dependent variable. She defined the three grading systems as: motor-outcome (100 percent on skill), managerial (100 percent on participation), and combination (50% skill, 30% cognitive, and 20% participation). This study was conducted in a field-based setting (nine riflery classes at a university activity program) instead of a laboratory and used an actual sport skill rather than an artificial task.

Three types of assessment instruments were used in this study: (a) a skill test (shooting scores), (b) a written test, and (c) a participation assessment. However, results on the written test and the participation assessment were not reported fully because the design of this study did not
plan to compare scores on the written test and participation assessment. For example, the written test was only taken by those students in the classes under the combinational grading system.

Although students under three different grading systems received different assessments and took different tests, standardized instruction was given across all classes under the three grading systems. Each of the nine rifle classes lasted ten weeks. The skill test consisted of a total of five shots after three warm-up shots. The same procedure was employed also in pretest, practice, and retention trial days.

According to the overall mean scores of skill tests, students under the combinational grading system scored highest (34.99) while the students with the participation grading system got the lowest points with the students of skill grading method in the middle position (33.47). However, the only statistically significant difference in the shooting scores was between the combination and the participation groups. She concluded:

If this study reflects the experience of many teachers and students, a teacher's choice of grading method would affect students' skill performances and learning (p. 48).

Furthermore, an explanation of the linkage between grading and student motivation was provided by Mueller, et al. (1988). Their assumption is that grading puts stress on
students and the stress forces students to learn. They conducted an experimental study to test this assumption. Two classes of bowling in a basic instructional program at the University of Northern Colorado were selected to serve as subjects. One was a control group and the other was an experimental one. The treatment group was graded in terms of skill improvement. In contrast, the control group was graded on attendance without regard to skill improvement. Instruction and student practice time which consisted of fourteen 50 minute class periods for eight weeks were kept consistent across the two classes. Although there were no significant differences between the two classes in terms of their adjusted post-test scores, they reported that the average adjusted post-test scores for the treatment group was found to be 4.95 pins higher than the adjusted post-test scores for the non-treatment group. They concluded:

The results of this study would indicate that more students improved their bowling averages and by a greater amount when grading was dependent upon required improvement .... Grade-induced stress would seem to facilitate skill improvement (p. 32).

While, some researchers believe that grade-induced stress may have negative effects on student outcome as a result of student fear of not performing well and being embarrassed (Goldstein & Tilkei, 1971; Hardy, 1974; Nixon & Jewett, 1980), other researchers showed that grade-induced stress had positive effects on student motor learning.
(Boyce, 1990; Klavora, 1979; Mueller, et al., 1988). In addition, Wrisberg (1983; 1984) and Nideffer (1979) reported that stress may facilitate or impair performance depending on the individual's response to stress and the nature of the motor task. Nideffer (1979) indicated that an appropriate grading system would bring an optimal level of stress for each individual and each motor task in controlling attention and concentration levels necessary for utmost proficiency.

**Grading Foci (Grading components)**

It is obvious that grading components are the determining factors for an appropriate grading system. It seems that grading foci become the guides and directions of student studying efforts given grading as a driving force in learning as shown in the above literature. Grading foci refer to those student learned capabilities such as motor skills, intellectual skills, cognitive strategies, attitudes, and verbal information (Dick and Carey, 1990; Gagne, et al., 1988) on which students are graded.

It is widely accepted that grading foci should follow the objectives of the instruction (Boyce, 1990; Dick & Carey, 1990; Gagne, et al., 1988; Harrison, 1983; Safrit, 1973). However, it also seems true that the grading foci become functionally the actual objectives of instruction. In other words, the content of what the grading system focuses on becomes the functional curricula. What students
learn is not necessarily what the teacher taught but what the grading system focused on (Son, 1989; Safrit, 1973; Toussignant, 1982). Therefore, grading foci merit special attention. The grading foci in physical education are well documented.

Theoretically, physical education affects all possible capability domains mentioned earlier; motor skill, intellectual skills, cognitive strategies, verbal information, and attitudes (Gagne, Briggs, and Wager, 1988; Safrit, 1973). The notion that physical education affects all of the human capability domains is well illustrated in the grading foci.

Some researchers (Fox, 1959; Hensley, et.al, 1989; Toussignant, 1982; Son, 1989) indicated that fewer tests have been given to evaluate the degree to which student performance changes while the reality remains that many teachers still use subjective judgment of such things as effort, attendance, sportsmanship, and performance to assign grades and to ascertain the effectiveness of instructional programs. However, it should be noted that the major characteristic of grading foci in physical education is its variety. A selection of major evidence about this variety in the grading foci is briefly reviewed.
Imwold, Rider, and Johnson (1982) reported that slightly more than half of the Florida public school physical education teachers utilized skills tests to assess performance of their students, while fewer than 40% used knowledge tests.

Having surveyed the Physical Activity Programs in 251 four-year colleges and universities with an undergraduate enrollment of 5,000 or more, Miller, et al. (1989) reported:

The most frequently used form of student evaluation or grading criteria was a combination of motor skill tests and knowledge tests used in 82 percent of the institutions. Six percent of the reporting institutions used only motor skill tests to determine grades and four percent of the schools surveyed employed knowledge tests only. Other criteria listed by respondents in determining grades were attendance, participation, written exams, and fitness tests. Most schools surveyed also noted that the grading criteria and the percent of such criteria is determined by the instructor (pp. 21-22).

Based on a survey of 1,396 returned questionnaires from high or middle schools physical education teachers in Iowa, Wyoming, Kansas, and Georgia; Hensley, Lambert, Baumgartner, and Stillwell (1987) reported that the grading focus of these teachers was related to attendance, attitude, dressing out, effort, homework, improvement, knowledge tests, participation, potential, subjective skills evaluation, skill tests, and sportsmanship. Overall, slightly less than half of the respondents reported regular use of skills tests and written examinations. Over half of the teachers
reported that the grading foci used were determined by the individual teacher. When asked about their use of subjective ratings to assess students, 33 percent indicated that they used subjective ratings frequently, while another 55 percent reported using subjective ratings occasionally. Participation was the most frequently used factor in determining student grades being selected by 96 percent of the respondents. Forty six percent of the respondents reported that participation was used to determine students' final grades. In addition, attitude accounted for 16 percent, followed by skill tests at 11 percent and attendance at 7 percent.

Bayless (1978) conducted an interesting survey comparing the opinions on grading practices as reported by 150 physical education teachers and 350 former students throughout Oklahoma. He reported:

The breakdown according to types of examination resulted in the following percentages: written examinations covering sports skills (15%); written examinations covering rules and regulations only (6%); written examinations covering first-aid skills (6%); written examinations covering health knowledge (11%); and combinations of sports knowledge, rules, and regulations (23%). These percentages are those of the professionals who claimed positive attempts in evaluation. The former students completing the questionnaire were very much in conflict in that they stated few written examinations were given in physical education. They claimed that most of their grades were derived from posture evaluation (10%) and physical fitness tests (31%) (p. 54-55).
A look at earlier reportings showed that students' grades in physical education were based solely on being present and in uniform daily (Mathews, 1978). A few years later Fox (1959) reported grading practices according to the percentage of responding physical education teachers in Oregon: personal hygiene, 98%; attitude, 96%; attendance, 94%; knowledge, 90%; achievement, 88%; physical development, 58%; improvement, 4%; and participation, 4%. A similar report 13 years later showed that physical education teachers in Louisiana public schools in 1972 (Coker) employed attendance, dressing out, and subjective skill evaluation as major foci.

Based on the responses of about 17 percent of the Idaho physical education teacher population, MacFarlane (1981) reported that consideration for awarding grades according to the percentage of teacher choices were ranked: student effort (98.3%), performance improvement (the difference between a pretest and a posttest) (70.3%). The next three components were, student rank in class (18.7%), student performance based on national tests or norms (23.4%), and the results of contract grading based on sets of performance objectives (30.7%). Also, he reported the common factors and the percentages of physical education teachers in Idaho used for grading were related to: neuromuscular skill development or skills achieved (17.6%), organic development
or fitness status or improvement (13.2), interpretation or knowledge (17.4), social-development (14.4), and administrative or managerial (37.2).

Pangrazi and Darst (1985) interpreted different grading foci or components as different viewpoints of evaluation. These different viewpoints were related to bivariate concepts; educational objectives (student outcomes) versus administrative tasks, process measures versus product measures, relative improvement versus potential, negative (deductive) versus positive (additive), and pass-fail versus letter grades. Furthermore, they describe five grading systems: (a) subjective observation systems in which a teacher makes a subjective judgment about a student performance through observation, (b) personal interviews in which the teacher sits down with individual students and questions their knowledge, as well as monitoring their skill levels, (c) self-evaluation, (d) contract grading system which is a student-paced approach to grading that allows students to progress at differing rates, and (e) global grading system in which all major components such as attitude, physical skill, physical fitness, and knowledge, are included.

Based on the previous grading practice literature, it seems that grading foci attend to major components; managerial factors, motor skill acquisition, and cognitive
learning. According to the weights of these three components, three grading systems for the purpose of research have been framed; managerial grading system, motor-outcome grading system, and typical grading system or combination system. Boyce (1990) defined three grading systems as: motor-outcome (100 percent on skill), managerial (100 percent on participation), and combination (50% skill, 30% cognitive, and 20% participation).

It should be noted that, apart from the research done by Boyce (1990) and Mueller, et al. (1988), there is a dearth of research findings to justify use of different grading criteria or grading components.

Gap Between Theory and Practice in Grading

A large gap exists between theorists, grading specialists, and practitioners (Wood and Safrait, 1990; Dunham, 1986). First, it seems reasonable to track the differences in the objectives of physical education to teacher planning (Placek, 1983). The objectives of physical education teachers focus on student behavior, class organization, and class management while theory suggests emphasizing student achievement or student products (Arrighi and Young, 1987; Goc-Karp and Zakrajsek, 1987; Placek, 1983; 1984; Stroot and Morton, 1989). Second, the attention of measurement specialists is directed at grading techniques such as test construction, grading validity, grading
reliability, and grading generalizability while practitioners are looking for practical solutions to their grading needs (Wood and Safrait, 1990). Third, theorists (Harrison, 1983) believe that grades should be based on achievement of all of the objectives of physical education, such as skills, physical fitness, knowledge, social skills, and attitudes and appreciations while teachers and practitioners are likely to select some aspects of the objectives of physical education to grade with little concern on the degree to which achievement has been obtained (Wood and Safrait, 1990). Fabricius, Hanson, Singer, and Solley (1967) stated:

The abyss between the theoretical and the practical is wide and deep as far as marking and grading is concerned and far too often the two never converge into a usable system....The theorist fails to give attention to obstacles universally found at the local school level. The local teacher, on the other hand, accepting limitations as irrevocable, has rejected the grading systems of the theorist, the result is compromise, mediocrity, and an inadequate job of evaluating and reporting pupil progress and achievement....(p. 35).

One of the explanations for the gap between theory and practice in grading is related to difficulties in grading that teachers face in daily life.

The difficulties in implementing grading in the real world are multiple dimensions such as limited time, insufficient equipment, overcrowded classes, teacher workload, lack of collegial and administrative support, and
differing teachers' beliefs about grading (Hensley, 1990; Hensley, Morrow, and East, 1990; Hensley, et al., 1987; Imwold, Rider, Johnson, 1982; King, 1990; Veal, 1990;). King (1990) pointed out that the misdirection given by theory and evaluation specialists in grading practice for practitioners makes it difficult to implement although the teacher still stands ready to receive the guidance they so well deserve and need. Another explanation for this gap between grading theory and practice is the fact that psychomotor measurement and evaluation is a growing area of study. Much of the research in psychomotor measurement and evaluation has focused on and will continue to focus on theoretical issues and the development of new methodologies (Wood & Safrait, 1990).

Some Other Issues Related to Grading in Physical Education

Grading methods in physical education activity courses have remained relatively unchanged over the past 30 years (Lumpkin & Avery, 1986; Miller, et al., 1989; Trimble & Hensley, 1984). According to these researchers, about 60-80% of physical education activity courses used letter grades, compared with about 30% who used satisfactory/unsatisfactory (or pass/non-pass) while a very small portion of the physical education activity courses employ other types of assessment such as a numerical
ratings. According to Lumpkin and Avery (1986), only 22% of the 2567 surveyed students preferred to have their physical education activity courses as elective while others prefer to have them required.

"Who should establish grading procedures?" is another important issue which is related to grading in physical education practice. It is reported that the important grading procedures are mainly developed and determined by teachers (Hensley, et al., 1989; MacFarlane, 1981). Hensley, et al. (1989), reported that grading procedures used by 54% of physical education teachers surveyed are developed by the teachers rather than by the school district or school administration. In MacFarlane's survey (1981), 81 percent of the physical education teachers indicated that they have some or total authority for grading procedures, while 19 percent used district wide criteria for grading.

Interestingly, it was found that student grades in college physical education activity classes are not necessarily solely determined by criteria such as student achievement. Instead, student grades are found to be correlated significantly with some characteristics of graduate teaching assistants in physical education (Ford, Puckett, & Tucker, 1987). After collection and analysis of data from 20 teaching assistants over two years on demographic data, teaching experience, GRE scores, total
score on the Tennessee Self-concept Scale, and teachers' ratings in conjunction with the average grade assignments, they concluded:

The present results suggest that assistants who teach activity courses in physical education assign grades in a fairly predictable manner. To help control the inflation of grades, it may be wise to train and monitor more closely assistants who appear to have proclivities to giving inflated grades, particularly women, single students, and those with relatively high self-concepts (pp. 738–739).

Summary of Chapter

The three sections in this chapter were: the important assumptions and theories of research in teaching as it relates to grading, the related research paradigms in teaching and in teaching in physical education, and research on grading in physical education.

1. Student achievement has been taken as a dependent variable. Student achievement could be categorized into intellectual skills, cognitive strategies, verbal information, attitudes, and motor skills. It is desirable to sample student achievement in all possible domains by using both process and product measures.

2. According to Theory X about motivation, the nature of a student is one which avoids responsibility and study unless he/she is controlled, directed, and driven.

3. Dunkin and Biddle provide a useful conceptual framework in which student achievement is causally related to other
independent variables of which grading could be considered an important one.

4. According to the process-product model, those variables related with teaching are used as independent variables while those variables related to student achievement are considered dependent variables. A large number of independent variables have been employed, however, a grading system as an independent variable has not been found in above process-product literature.

5. The ecological paradigm emphasizes the complexities of a natural instructional environment. Task structures are important conceptual frameworks for understanding these complexities. Accountability systems are integral parts in these conceptual frameworks of task structures.

6. ETU studies represent the final step of the descriptive-correlational-experimental loop in research on teaching in physical education.

7. Grading systems have been considered an essential and powerful variable in driving, mediating, and directing student learning. Great variety is found in grading systems and grading components. It seems that there is a lack of knowledge about the nature of the grading system as a driving and motivating force. Furthermore, almost all findings on grading systems have been descriptive. Knowledge on the effects of grading systems on student
learning is still limited primarily to some qualitative descriptions. Only two experimental studies on the effects and the mechanics of grading systems in terms of the roles of independent variable on student achievement have been reported. Also, research questions related to grading systems, which could be most appropriately addressed by experimental methodology, have been raised but not answered.
CHAPTER III
METHODOLOGY

The purpose of this study was to test and identify the effects of three grading systems on student outcome measures in a basic instruction program. A quasiexperimental design (Kennedy & Bush, 1985; Shavelson, 1981) study was used wherein each intact class was randomly assigned to a treatment condition during Spring Quarter, 1991 at The Ohio State University. Subjects employed in this experimental study were students in three table tennis classes.

Procedures

This quasiexperimental study had three phase; pre-treatment, treatment, and post-treatment.

During the pre-treatment phase, a pilot study was conducted with three similar table tennis classes during Winter Quarter, 1991. The purposes of the pilot study were to:

1. Test the reliability of the two skills tests: one-minute volley test and serve test.

2. Develop attitude items for the questionnaire and test for reliability.
3. Select a process measurement instrument and practice using it.

4. Select a teacher behavior congruence instrument and practice using it.

5. Practice working with videotape and audiotape devices.

Also, in the pre-treatment phase, human subject clearance was obtained (Appendix A), course outlines, unit teaching plans, and a first week lesson plan for all three classes were developed (Appendix B). These unit teaching plans and the first week lesson plan were followed in the treatment and the post-treatment phases.

Pretests (written exam and skill tests) were administrated before introducing the unit of instruction in each table tennis class. All of the pretests were completed before students were informed about the grading system which would be used in their class. Pretests were conducted during the first three meetings.

Students in all three classes were briefly informed at the beginning of the first meeting that they were involved in a research study. They were told that information collected from their tests would be useful to students in future physical education classes.

During the pre-treatment phase, students were allowed to add or change sections.
The treatment phase started with an introduction of the course outline which included a description of the grading system. A verbal explanation detailing the grading system was given along with other course information. Prior to the start of tournament play and the post-tests, the grading systems were reviewed verbally for each class.

After introducing the grading system, no student was allowed to add or change sections.

Normal instruction started on the fourth day (the first three meetings were spent on pretesting) and was guided by the unit teaching plan developed in the pre-treatment phase. In addition, notes were taken daily during the first lesson on details such as timing of transitions, feedback, questioning, and praising. These notes were used to guide the next two classes. Furthermore, a videotape machine with a wide angle lens was set up in a corner of the gymnasium where it was turned on before the start of a lesson and turned off after the end of the lesson. After classes were over each day, the investigator obtained information from playback of the videotape. This information worked as daily feedback to ensure that the investigator's behaviors were consistent across the three classes.

The major purpose of videotaping in this phase was to provide objective and reliable data on the investigator's teaching behaviors across the three classes in terms of
consistency. After reliability between the investigator and another coder (Coder "J") reached 90% for two consecutive four-minute training sessions, Coder "J" was allowed to start coding.

The first half of the treatment phase was spent in covering basic table tennis skills and the next half was spent in game play. Measures on game play variables were collected daily by students, and attendance rates were recorded by the investigator daily.

Within the post-treatment phase, posttests on two skill tests, a written test, and a 15-item questionnaire were administered. Three days were spent in the post-treatment phase.

After the raw data on the posttests, game play variables, and attitude questionnaire were collected; they were entered together with those data collected in the pre-treatment and the treatment phases into the computer and analyzed through SAS (Statistic Analytic Systems).

Subjects

Students in three table tennis classes in the basic instruction program at the Ohio State University in Spring Quarter, 1991 comprised the subjects. There were 14 students in each of the classes under the Skill grading system and the Managerial grading system and 16 students in the class under the Typical grading system. Class rank of
these students included freshmen to graduate students. Most of the students had some experience playing table tennis.

Although table tennis is not popular in this country, it is similar to other racket sports such as tennis, badminton, and racquetball which are more popular. Table tennis requires a small practice area and provides ideal conditions for experimental control, video taping, and observation.

The meeting place for all three classes was a spacious and appropriately illuminated multipurpose room located in Larkins Hall. There was one table for every two students. A gymnasium aide set tables up on each meeting day. Paddles and balls were provided by the program area.

Design Model

This study utilized a quasiexperimental design with three groups as shown in the following model:

<table>
<thead>
<tr>
<th>Groups</th>
<th>pretests</th>
<th>treatments</th>
<th>posttests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x</td>
<td>Skill</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td>Managerial</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>x</td>
<td>Typical</td>
<td>x</td>
</tr>
</tbody>
</table>

Figure 1 Illustration of Design Model
Three table tennis classes assumed the roles of the three groups. The three treatment groups were subjected to one of the three grading systems based on practical and theoretical knowledge. Each intact class was randomly assigned to one of the three grading systems. Before and after implementing treatments there were pretests and posttests.

Possible Threats to Internal Validity

Because of the nature of the design, some possible threats to the internal validity in this study were raised. These possible threats to internal validity included; variation of selection, researcher bias, and loss of student cooperation. In addition, the relative methods against these possible threats were employed.

Variation of Selection

Since this was a quasiexperimental design and the data measurement was based on individual students, the following two threats were likely. First, differences in the dependent variables might have existed before implementation of treatments; second, "blocking" effects might have played a role. To keep the selection threat to a minimum, the researcher first used adjusted data and secondly assumed that the "blocking factor" was at a level which was minimized as much as possible. Many quasiexperimental research studies conducted in education do this because
random assignment of experimental units to treatments is ethically inappropriate.

**Researcher Bias**

In this study, the teacher for the three table tennis classes and the researcher were the same person. Knowing the purpose of the research, there was a threat to the teaching behaviors or other variables from possible reactions of the researcher. For example, the researcher might teach more effectively in the class utilizing the Skill grading system while managing better in the Managerial grading system class.

To control for the threat of researcher reactivity, all three classes were videotaped at least once weekly throughout the quarter during which this experiment was conducted. Congruence data on the teacher's behaviors were collected and analyzed to provide feedback and information for monitoring and correcting teacher behaviors.

Congruence data were captured by using the Direct Instruction Behavior Analysis (DIBA) (Zakrajsek and Tannehill, 1989). DIBA was chosen because: (a) data collected from DIBA categories of teacher behaviors were sufficient to provide information to tell whether the teacher's behavior was consistent among the three treatments, (b) the teacher employed a direct teaching style in all three classes since the teacher was more comfortable
with direct teaching, (c) data from student behavior categories of DIBA met the needs of process measures on the dependent variables, and (d) DIBA is relatively easy to use in terms of training coders and coding (for detail, refer to Appendix C).

**Loss of Student Cooperation**

Since the two grading systems; Managerial and Skill, represent two extremes, some problems might arise related to student cooperation during the period of treatment. For example, in the class treated by the Skill grading system, students might have behaved inappropriately to express their dissatisfaction with the grading system. In contrast, in the Managerial grading system class, students might have been too relaxed in terms of skill learning.

**Treatments or Independent Variables**

Three different grading systems, independent variables, are shown in Table 1 where "S" represents the Skill grading system, "M" is the Managerial, and "T" refers to the Typical grading system.

Participation and attendance consisted of the following components; being physically present, dressing appropriately, being punctual, staying until class was over, being on task, and others.
Table 1  Three grading systems used as independent variables

<table>
<thead>
<tr>
<th>Components</th>
<th>S</th>
<th>M</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation/attendance</td>
<td>0</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>Skill tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Serve test</td>
<td>25</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>(b) One minute volley test</td>
<td>25</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>(c) Game play variables</td>
<td>50</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Written Exam</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>

It should be noted that the derivation of the three grading systems was based on theory and practice. These three grading systems represented three common grading systems and maximized the possibility of showing different effects on measures of student achievement.

The Skill grading system was primarily derived from theory (Boyce, 1990) for the purpose of this study. Probably, no physical education teacher uses a grading system in which students are graded 100% on skill tests.

The Managerial grading system is widely used in public school physical education (Coker, 1972; Mathews, 1978; Son, 1989; Tousignant, 1982) and in some Basic Instructional
Programs at colleges and universities (Miller, et.al, 1989). However, the most prevalent grading system used in Basic Instruction Programs at colleges and universities is a grading system with multiple components in which students are graded by several categories; skill tests, written tests, participation, and effort (Miller, et.al, 1989). This grading system with multiple components is called the Typical grading system.

It should be noted that verification of selection of the above three grading systems as independent variables was strengthened by Boyce's (1990) recent study in which she employed three similar grading systems as independent variables.

Furthermore, a review of the grading systems used in the Basic Instructional Program at the Ohio State University, also showed the likelihood of the validity of the three grading systems. Percentage of grading systems used in the Basic Instruction Program at the Ohio State University for sports requiring the use of a "racket" (table tennis, badminton, squash, tennis, and racquetball) for the previous two years is shown in Table 2.
Table 2
Grading systems for racket sports at OSU (%)

<table>
<thead>
<tr>
<th>Components</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part./Attend</td>
<td>22</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>Skill test</td>
<td>30</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Written test</td>
<td>32</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Game</td>
<td>14</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Others</td>
<td>4</td>
<td>40</td>
<td>0</td>
</tr>
</tbody>
</table>

By combining Participation/Attendance with Game in Table 1, this category accounted for 36%. It is obvious that an average instructor for the listed courses used a grading system which had multiple components. However, there was widespread variation in the sense that some teachers used a managerial-like grading system while others used an outcome-like grading system.

It should be noted that in this experimental study, other factors and variables aside from these three grading systems were kept the same as much as possible.

Duration of Treatment

Duration of treatment referred to the treatment phase.

The study was conducted Spring Quarter, 1991 which lasted 10
weeks. Students met twice a week. One holiday took one of the 20 meeting days. Three lessons at the beginning and end of the quarter were spent in pre-tests and post-tests to collect data on skill tests, written test, and attitude questionnaire. Therefore, there were only 13 actual lessons of instruction and game play. Each lesson lasted about 35 minutes with seven minutes taken for students to shower and change before and after the lesson.

Dependent Variables

The dependent variables were student outcome measures which related to; student behaviors, attendance rates, one-minute volley test, serve test, written test, variables on game play, and an attitude questionnaire.

Student Behaviors

Measures on six student behaviors were taken; motor engaged, cognitive engaged, response preparing, getting equipment or relocating, waiting, and off-task, which comprised the student behaviors of the Direct Instruction Behavior Analysis system (Zakrajsek and Tannehill, 1989) (Appendix C).

Data were entered through playback of videotapes. There were two coders; coder "M" (who was the investigator) and coder "J". After reaching a 90% agreement rate between the two coders for two consecutive four-minute intervals, coding was allowed. The agreement criterion was set at 85%
in advance. Coder "M" coded a total of eight lessons for all three classes while coder "J" coded three.

**Attendance Rates**

Attendance rates, which was the percentage of students who physically attended classes each day, were collected by the investigator.

**One-Minute Volley Test**

This table-tennis test was developed by Mott and Lockhart (1946). For this test in this study, a table-tennis table is needed and a post board of similar material as the table surface. This board is placed perpendicular to the playing surface in the middle of the table (between the two sides), to serve as a backboard. The net is set according to the Rules so that it is close to the perpendicular board. Extra balls are placed at the side of the testing table. A stopwatch, table-tennis paddle, and three table-tennis balls are also necessary.

On the signal "Go", the player drops a ball to the table and volleys it against the perpendicular board as many times as possible in one minute. Any number of bounces on the playing surface is permitted. Hits on the perpendicular surface do not count if the ball is volleyed (that is, the ball must bounce at least once), the player puts the free hand on the table during or immediately preceding a hit or after the ball strikes the perpendicular
surface on the net. The test score is the mean score of two trials.

According to the results of the pilot study, the reliability coefficients for the one-minute volley test were 85.8% with minimum value 64.1% and maximum value 97.8%. The reliability coefficient was the mean of differences between the first and second tests for each student. This was lower than the reliability coefficient of 0.90 with college women as reported by Mott and Lockhart (1946). Mott and Lockhart also reported that a validity coefficient of 0.84 was obtained, but the criterion was not given. Unfortunately, during this pilot study an appropriate criterion for establishing validity was not found.

Serve Test

Each student was required to demonstrate both forehand and backhand (defined by skill forms,) serving skills by serving the balls so that each ball lands on any arranged quarter of the opponent's side of the table surface. Each student served ten trials for the forehand and ten trials for the backhand. Each serve was judged by the official rules of table tennis. Each good serve was worth 0.5 point. Each good serve landing within an arranged quarter counted one point. Each unsuccessful serve was given zero points. Total score of twenty trials was the final score for this serve test. The reliability coefficient for the serve test
was 91.9% with a minimum 70.6 and maximum 100.0% according to results from the pilot study.

Written Examination

The written examination was to test cognitive achievement related to playing table tennis. This written examination consisted of 10 true/false questions and 10 multiple choice questions. These questions were based on the course objectives and content covered. In other words, these questions were related to those student capabilities (verbal information, intellectual skills, and cognitive knowledge), which are essential to playing table tennis (for detail, refer to Appendix D).

Skills Demonstrated on Game Play Variables

In this part of the skill test, the results of each return, forehand serve, and backhand serve of game play were recorded. Results of the forehand serves and backhand serves were categorized as good serves, serves scored, and serves missed. The number of returns in each rally and the number of rallies in each game were categorized as returns scored, rallies scored, returns missed, and rallies missed.

Students under the Skill grading system and the Typical grading system were informed as follows:

You will be graded in terms of the quantity and quality of your game play. Quantity refers to the numbers of games, volleys, serves, and returns. The more the games, rallies, serves, and returns you do, the more practice you have, the higher your grade should be.
Quality refers to the number of games won, returns scored, and serves scored, etc. The number of games won, returns scored, and serves scored will be positively related with your grade while the number of games lost, returns missed, and serves missed will be negatively related with your grade.

Students under the Managerial Grading System were informed in a similar way that they should concentrate on both quantity and quality in game play. However, they were reassured that their grades would not be related to the results of game play.

Before the start of the tournament, one lesson was spent in explaining the definitions for each category of the game play variables and training students to keep game play scores. Definitions of game play variables are listed in Appendix E.

**Game play format and organization.** Students were grouped in fours according to skill levels. Two students played at a time while the other two were recording. Each class was divided into two parts. During the second part, the two players and the two recorders switched roles. The four students played one by one. It took three meetings for one of the four to finish competition against the other three.

**Procedure for game play recording.**

1. Students were grouped into fours. Two students played while the other two recorded.
2. Students were given time to get familiar with each category of definitions, although the categories in this "GAME PLAY RECORDING SHEET" (Figure 2) were similar to those concepts used in keeping score in playing a table tennis game.

3. Recorders did not make marks until the rally was over. No player started the next serve until the recording had been completed. Only serves or returns which resulted in a point were recorded.

4. In marking a serve, the recorder needed to determine which category the serve was and then wait until the rally was over to record it.

5. In recording a rally, the number of returns the target player performed was recorded. The number was written under "Scored" column if the target student won that point and under "Missed" column if the target student lost the rally. Figure 3 shows an example.

Reliability check. Total number of rallies of the two recorders should equal the total number of serves from the two recorders. Number of returns in each rally between the two recorders must be no more than one because a serve which started a rally was not counted as a return.
GAME PLAY RECORDING SHEET

Name of player (last) __________________ (first) ____________
Name of opponent (last) ____________ (first) ____________
Name of recorder (last) ____________ (first) ____________

Class time (Check one): 1. 12:00 M W
2. 1:00 M W
3. 12:00 T R
Amount of game time: ______________ minutes.

<table>
<thead>
<tr>
<th>Good Serves</th>
<th>Scored</th>
<th>Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOREHAND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>total</td>
<td>total</td>
</tr>
<tr>
<td>BACKHAND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>total</td>
<td>total</td>
</tr>
</tbody>
</table>

* * * * * * * * * *

<table>
<thead>
<tr>
<th>Scored</th>
<th>Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RALLY</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># of Total Return</td>
<td># of Total Return</td>
<td></td>
</tr>
<tr>
<td># of Total Rally</td>
<td># of Total Rally</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GAME</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>SCORES</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>LOST/WON</th>
<th></th>
</tr>
</thead>
</table>

Figure 2  Game Play Recording Sheet
<table>
<thead>
<tr>
<th>Scored</th>
<th>Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, 2, 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>RALLY</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td># of Total Return</td>
<td># of Total Return</td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
<tr>
<td># of Total Rally</td>
<td># of Total Rally</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Note: This target student played three rallies. For the first rally, she returned the ball three times. For the second one, she returned twice, and for the last rally, she returned the ball seven times. She won all three rallies.

Figure 3 Example of Game Play Recording and Computation

Measures on Student Attitude Questionnaire

Measures on student attitude toward the three grading systems were obtained through administration of a 15-item questionnaire. Because a compatible pre-test on this questionnaire was not available this questionnaire was administered once during the post-treatment phase. It was assumed that students in all three classes had similar attitudes toward physical education before they took one of these table tennis courses. The 15-item questionnaire is presented in Appendix F.

The reliability coefficients for those 15 questions according to the results of the pilot study were (a) >.90
for questions 1, 2, 3, 5, 6, 8; (b) >.80 and <.90 for 4, 9, 10, 11, 12, 13, 14; and (c) = 0.62 for question 7.

Data Analysis

The primary purpose of data analysis in this study was to obtain statistics from the samples and make inferences from the statistics to the parameters of the population. Several tests were used.

1. Analysis of covariance (ANCOVA) was employed in analyzing the differences of pre- and posttest data among the three treatment classes. These data were related to the one-minute volley test, the serve test, and the written examination.

2. Analysis of Variance (ANOVA) was used in analyzing the differences among the three groups (the Skill, the Managerial, and the Typical) on game play variables, teacher behaviors, and student behaviors which had no pretests scores.

3. Nonparametric statistical analyses were used where the nonparametric analysis was obviously more appropriate. Since the data derived from the student attitude measures were ranked-data and the data of student attendance rates were not appropriate for parametric analysis, the nonparametric statistic analysis was employed. All nonparametric statistical analyses which were available on Statistic Analytic System (SAS) were run on the data. These
analyses were: Wilcoxon Scores, Kruskal-Wallis Test (Chi-Square Approximation), Median Scores, Median 1-Way Analysis, Van der Waerden Scores and Van der Waerden 1-Way. Since there was no obvious advantage for any specific analysis, the results from the Kruskal-Wallis Test (Chi-Square Approximation) was chosen to present the results.

In order to increase the power or efficiency of data analysis on the attitude measures, the Principle Component Analysis was used.

4. All data were processed through SAS.
CHAPTER IV

RESULTS

The purpose of this study was to identify and test the effects of three grading systems on student outcomes in a basic instruction program. This chapter presents results of parametric and non-parametric statistical procedures on the measures of student outcomes. The analyses test the null hypotheses stated in Chapter 1.

The organization of this chapter first verifies statistically the consistency of teacher behavior across treatment classes. Having established teacher consistency, comparative analyses are presented on student behaviors, attendance rates, skill tests, written tests, game play variables, and attitudes among classes of students under three grading conditions.

**Teacher Behavior Consistency Across the Three Treatment Classes**

The teacher for the three table tennis classes under the three treatment conditions was the investigator. To keep researcher bias at a minimum, several approaches were
employed. First, a daily lesson plan for the unit was
developed prior to beginning the study (Appendix B.).
Second, brief notes were made in the first class on
important teaching events such as the timing, direction,
duration, and times of transitions and these notes were used
to guide the following two classes. The final method to
ensure that the teacher's behavior was consistent among the
three classes was to videotape at least one lesson per week,
code the tapes, and submit the data for statistical
Instructional Behavior Analysis (DIBA) was used to code
teacher behavior data which was entered by two coders. The
investigator was labelled coder "M" and another independent
coder was coder "J". Data on teacher behaviors derived from
each coder are presented separately in Table 3 and Table 4.

Of the 11 teacher behaviors, the first eight are the
major categories all of which are exclusive. The last three
(enthusiasm, clarity, and modeling) are subscripts for the
eight major categories. In other words, they are only coded
in conjunction with one of the first eight and can be
partialled out for separate analysis.
Table 3. DIBA Data on Teacher Behaviors

(Coder J)

<table>
<thead>
<tr>
<th>Teacher Behaviors</th>
<th>Skill</th>
<th></th>
<th>Managerial</th>
<th></th>
<th>Typical</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Informing</td>
<td>24.00</td>
<td>17.69</td>
<td>20.33</td>
<td>15.04</td>
<td>24.67</td>
<td>19.01</td>
</tr>
<tr>
<td>Observing</td>
<td>75.33</td>
<td>40.00</td>
<td>79.00</td>
<td>32.19</td>
<td>74.00</td>
<td>33.00</td>
</tr>
<tr>
<td>Structuring</td>
<td>12.67</td>
<td>2.52</td>
<td>14.00</td>
<td>1.73</td>
<td>18.67</td>
<td>1.15</td>
</tr>
<tr>
<td>Questioning</td>
<td>4.67</td>
<td>2.31</td>
<td>2.67</td>
<td>2.08</td>
<td>3.67</td>
<td>0.58</td>
</tr>
<tr>
<td>Praising</td>
<td>8.67</td>
<td>9.29</td>
<td>13.67</td>
<td>15.18</td>
<td>7.67</td>
<td>4.93</td>
</tr>
<tr>
<td>Feedback</td>
<td>8.33</td>
<td>6.81</td>
<td>8.67</td>
<td>4.93</td>
<td>7.67</td>
<td>4.04</td>
</tr>
<tr>
<td>Controlling</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.33</td>
<td>1.53</td>
</tr>
<tr>
<td>Non Above</td>
<td>9.67</td>
<td>9.18</td>
<td>6.33</td>
<td>3.21</td>
<td>11.00</td>
<td>6.24</td>
</tr>
<tr>
<td>Enthusiasm</td>
<td>0.00</td>
<td>0.00</td>
<td>0.67</td>
<td>1.15</td>
<td>1.00</td>
<td>1.73</td>
</tr>
<tr>
<td>Clarity</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.67</td>
<td>2.89</td>
</tr>
<tr>
<td>Modeling</td>
<td>7.33</td>
<td>8.08</td>
<td>7.00</td>
<td>7.00</td>
<td>11.00</td>
<td>12.77</td>
</tr>
</tbody>
</table>
Table 4. DIBA Data on Teacher Behaviors

(Coder M)

<table>
<thead>
<tr>
<th>Teacher Behaviors</th>
<th>Skill Mean</th>
<th>Skill SD</th>
<th>Managerial Mean</th>
<th>Managerial SD</th>
<th>Typical Mean</th>
<th>Typical SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informing</td>
<td>29.12</td>
<td>19.38</td>
<td>26.38</td>
<td>17.45</td>
<td>29.38</td>
<td>18.48</td>
</tr>
<tr>
<td>Observing</td>
<td>74.38</td>
<td>33.59</td>
<td>77.25</td>
<td>30.66</td>
<td>75.75</td>
<td>28.10</td>
</tr>
<tr>
<td>Structuring</td>
<td>18.50</td>
<td>7.45</td>
<td>18.63</td>
<td>8.23</td>
<td>20.13</td>
<td>5.57</td>
</tr>
<tr>
<td>Questioning</td>
<td>4.38</td>
<td>1.41</td>
<td>3.50</td>
<td>1.85</td>
<td>5.13</td>
<td>2.53</td>
</tr>
<tr>
<td>Praising</td>
<td>7.13</td>
<td>5.49</td>
<td>8.75</td>
<td>7.70</td>
<td>6.63</td>
<td>3.38</td>
</tr>
<tr>
<td>Feedback</td>
<td>6.75</td>
<td>4.59</td>
<td>6.25</td>
<td>5.55</td>
<td>5.25</td>
<td>3.11</td>
</tr>
<tr>
<td>Controlling</td>
<td>0.13</td>
<td>0.35</td>
<td>0.88</td>
<td>1.13</td>
<td>0.38</td>
<td>0.52</td>
</tr>
<tr>
<td>Non Above</td>
<td>7.13</td>
<td>6.62</td>
<td>6.13</td>
<td>4.45</td>
<td>7.00</td>
<td>6.28</td>
</tr>
<tr>
<td>Enthusiasm</td>
<td>0.50</td>
<td>1.07</td>
<td>0.88</td>
<td>1.46</td>
<td>0.63</td>
<td>1.41</td>
</tr>
<tr>
<td>Clarity</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.63</td>
<td>1.77</td>
</tr>
<tr>
<td>Modeling</td>
<td>6.38</td>
<td>5.97</td>
<td>5.50</td>
<td>5.68</td>
<td>7.63</td>
<td>8.25</td>
</tr>
</tbody>
</table>
A one-way analysis of variance showed no significant difference in any of the 11 teacher behaviors among the three classes at the .05 alpha level. Of note were some rather large standard deviations (SD) for most of the variables in Table 3. For example, for the first variable "informing", the standard deviations ranged between 15.04 and 19.01 with means ranging between 20.33 and 24.67. The large SD reflected the fact that the teaching tasks varied largely in classes observed. For example, "informing" scored higher when the major objective was to state the factors related with skill forms, spins, speeds, landings, and trajectories of both forehand sidespinning and underspinning serves and for the same reason, "informing" would score very low in sessions when the major activity was game play.

Both coder "J" and coder "M" provided similar data even though coder "J" observed three sessions and coder "M" observed eight sessions. Based on the analysis of variance it was concluded that the investigator's behaviors were consistent across the three treatment classes on the 11 teacher behaviors as defined by DIBA.

Comparisons of Grading Effects on Student Behaviors by DIBA Categories

The major purpose of this experimental study was to identify the effects of three different grading systems on
student achievement. As reviewed in Chapter 2, process measures for student achievement have been widely used in teacher effectiveness research and research on teaching in physical education. Process measures become realistic where systematic observation systems are available. In addition, in order to facilitate understanding the results derived from mean comparisons on student behaviors, it should be noted that time-interval observation technique plays an important role in most systematic observational systems (Darst, Zakrajsek, & Mancini, 1989). Collecting data on both teacher and student behaviors are possible using the Direct Instruction Observation Analysis (Zakrajsek & Tannehill, 1989), which is a three-second interval systematic observation instrument in which a time-interval technique plays a determinant role.

Data obtained through DIBA are presented in Table 5. There are six dimensions of student behavior: student motor engaged, cognitive engaged, response preparing, getting equipment or relocating, waiting for a turn, and off-task. Although there were numeric differences among the three classes under different treatment conditions, none of the behaviors tested significantly different among the three classes at the .05 alpha level. It should be noted that, for the student behavior "Off-Task", the total of sum squares of ANOVA was zero which the SAS program had
difficulty handling. As a result of the zero total of sum squares, the SAS program reported a high "F" (99999.99) and zero p-value which should be ignored in the sense that all the student "off-task" behavior across the three classes was zero. Furthermore, results of ANOVA showed that it was unlikely that any of these numerical differences were due to the treatments (Table 5) because there was not sufficient evidence to reject the null hypothesis that there would be no treatment effects on student process measures. The most likely explanation for the fact that the data from DIBA did not identify treatment effects on student process measures may be related to the nature of time-interval measures and the well-controlled instructional environment. The nature of time-interval measures implies that the "quality" of time such as activity rate or opportunity to respond (OTR) was ignored while the well-controlled instructional environment meant that each class under the three treatments was carefully organized to have the same amount of allocated time. Thus, it is more likely, that there were no significant differences in terms of student use of time according to the categories defined by DIBA.
Table 5. Comparison of Grading Effects on Student Behavior by DIBA Categories (Coder M)

**Dependent Variable: Student Motor Engaged**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>42.25</td>
<td>0.13</td>
<td>0.88</td>
</tr>
<tr>
<td>Residual Error</td>
<td>21</td>
<td>3498.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>23</td>
<td>3540.62</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Dependent Variable: Student Cognitive Engaged**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>3.00</td>
<td>0.01</td>
<td>0.99</td>
</tr>
<tr>
<td>Residual Error</td>
<td>21</td>
<td>4173.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>23</td>
<td>4176.62</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Dependent Variable: Student Response Preparing**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>0.33</td>
<td>0.07</td>
<td>0.93</td>
</tr>
<tr>
<td>Residual Error</td>
<td>21</td>
<td>49.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>23</td>
<td>49.96</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5 (Continued)

<table>
<thead>
<tr>
<th>Dependent Variable: Student Getting Equipment or Relocating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Classes</td>
</tr>
<tr>
<td>Residual Error</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent Variable: Student Waiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Classes</td>
</tr>
<tr>
<td>Residual Error</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent Variable: Student Off-Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Classes</td>
</tr>
<tr>
<td>Residual Error</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>
Comparison of Grading Effects on Student Attendance Rate

Attendance was defined in this study as physical presence in class. Because the scores for students "off-task" were zero for all three classes (Table 5) it seems reasonable to infer that attendance is similar to participation in the following discussion since all students were on-task if they showed up.

Because attendance is necessary for student learning in a physical education class, it seems imperative to collect data on attendance and participation rates and to report results derived from such an analysis.

According to Table 6, means of attendance rate for the Skill class, Managerial class, and Typical class were 85.71%, 97.80%, and 92.31%, respectively. Attendance rate was the percentage of the total number of registered students in each class in attendance during the quarter. "S" stands for the Skill grading system, "M" for the Managerial grading system, and "T" represents the Typical grading system. The Managerial class had a significantly higher attendance rate than the other two classes (Kruskal-Wallis Test of Chi-Square Approximation (df=2) = 14.86, p<.01. The Kruskal-Wallis Test statistic was 13.57 (df = 1, p<.01) for the comparison between the Skill class and the Managerial class, and it was 5.02 (df = 1, p<.03) between
the Skill and the Typical classes. In addition, it should be noted that the Chi-Square Approximation for the Kruskal-Wallis Test (df = 1) = 3.64 (p-value = .0565) between the Managerial and Typical classes was close to .05. Student attendance rates for a total of 13 lessons for each treatment class was used without inclusion of three pre-test and three post-test lessons.

### Table 6. Comparison of Grading Effects on Student Percent of Attendance Rate.

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>13</td>
<td>85.71</td>
<td>8.75</td>
<td>64.29</td>
<td>100.00</td>
<td>S-M  *</td>
</tr>
<tr>
<td>Managerial</td>
<td>13</td>
<td>97.80</td>
<td>3.43</td>
<td>92.86</td>
<td>100.00</td>
<td>S-T  *</td>
</tr>
<tr>
<td>Typical</td>
<td>13</td>
<td>92.31</td>
<td>6.53</td>
<td>82.35</td>
<td>100.00</td>
<td>M-T</td>
</tr>
</tbody>
</table>

* = p<.05

Attendance and participation components accounted for 100% of the grade for those students in the Managerial class and these students had the highest attendance rate. The same component explains 40% of the grade for students in the Typical class, and these students had the second highest attendance rates. As a result of the Skill grading system (students were only graded on skill tests), the Skill class had the poorest attendance rate. Cause and effect seems
evident between student attendance rates and the percentage of attendance explanation for grades. Based on the statistical analysis, it was concluded that the greater the use of attendance as a grading component, the greater the attendance rate of students.

**Comparison of Grading Effects on Skill Tests**

Executing basic serving and returning skills of table tennis were the major objectives for students in these three table tennis classes. Pre- and posttests were given on a one-minute volley and a serve test.

Results of an analysis of covariance (ANCOVA) on the one-minute volley tests are presented in Table 7. First, it is noted that the correlation coefficient for the pretest and posttest scores was .58. The second line of the ANCOVA table shows the part of total variance in posttest scores explained by the pretest scores. In other words, the total variance of the posttest scores was 11576.08 and the information in the relationship between the pretest and posttest scores explained 3896.61 out of the 11576.08 variance in the posttest scores. The philosophy here is that the relationship between the pretest and posttest provides more information about the variance in the posttest scores, therefore we are able to analyze the variance in posttest scores with less errors, hence, more accurately. In other words, ANOVA analyzes the variance in posttest
scores while taking the influence of pretest scores on posttest scores into consideration because of the correlation between the pretest and the posttest.

Table 7 shows that the treatment effects on the posttest scores were significant, F(2, 84) = 6.93, p<.01. Post hoc tests, mean comparisons, (Table 8) show that, means for the Skill class, the Typical class, and the Managerial class in order were 90.93, 88.28, and 81.93. The Managerial class scored significantly lower than both the Skill class and the Typical class, p<.05. There was no significant difference for this test between the Skill and the Typical class at the .05 level.

Similar results are shown in the test scores on the serve test (Tables 9 and 10). First, the treatment effects on the posttest scores among the three treatment classes were significantly different, F(2, 82) = 7.56, p<.01. Second, mean comparisons (Table 10) show that the score of the Managerial class was significantly lower than both the Skill class and the Typical class, p<.05. Similarly, there was no significant difference between the Skill class and the Typical class with alpha at .05. Based on the above analysis, the skill test scores were higher when related to grades and lower if skill tests were not related to grades. Therefore, the null hypothesis that three different grading system have no effects on the scores of skill tests was
rejected. However, it should be noted that the correlation coefficient was rather low between the pre- and the posttests, \( r = .27 \).

**Comparison of Grading Effects on Written Test**

Understanding the basic rules and knowledge on skills/strategies of playing table tennis by solving written questions was stated as a course objective for the three treatment classes. It is widely accepted that verbal information and capabilities which can be adequately reflected in written questions are essential or useful for accomplishing goals in physical education (Dick & Carey, 1990; Mao & Zakrajsek, 1988). Written Pre- and posttests were given to students in each class.

Of the three classes, only those students in the Typical class were graded in terms of their performance on the written examination. Table 11 shows the result of ANCOVA on the written scores. The three grading systems showed significant effects on the class scores for the written examination, \( F(2, 40) = 4.54, p < .01 \). Table 12 shows that the Typical class scores (Mean = 10.17) were significantly higher than both the Skill class (Mean = 6.13), \( p < .05 \), and the Managerial class (Mean = 5.73), \( p < .05 \). There was no significant difference between the Skill class
Table 7. ANCOVA On One-Minute Volley Tests

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to Regression</td>
<td>1</td>
<td>3896.61</td>
<td>49.65</td>
<td>.01</td>
</tr>
<tr>
<td>Post-Test</td>
<td>2</td>
<td>1087.08</td>
<td>6.93</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>84</td>
<td>6592.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>87</td>
<td>11576.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Mean Comparison On One-Minute Volley Test Scores

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean (Pre)</th>
<th>SD (Post)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>14</td>
<td>90.93 (75.28)</td>
<td>12.30 (16.37)</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>14</td>
<td>81.93 (70.82)</td>
<td>8.49 (10.32)</td>
<td>M-T *</td>
</tr>
<tr>
<td>Typical</td>
<td>16</td>
<td>88.28 (67.28)</td>
<td>11.84 (16.65)</td>
<td>S-T</td>
</tr>
</tbody>
</table>

* = p<.05
Table 9. ANCOVA Table On Serve Tests

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to Regression</td>
<td>1</td>
<td>4.61</td>
<td>7.58</td>
<td>.01</td>
</tr>
<tr>
<td>Post-Test</td>
<td>2</td>
<td>9.17</td>
<td>7.56</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>82</td>
<td>49.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>85</td>
<td>63.52</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10. Mean Comparison On Serve Tests.

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean(Pretest)</th>
<th>SD(Pretest)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>13</td>
<td>9.29 (8.25)</td>
<td>0.47 (1.09)</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>14</td>
<td>8.77 (8.48)</td>
<td>1.10 (1.39)</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>16</td>
<td>9.47 (8.08)</td>
<td>0.75 (1.50)</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

* = p<.05

r (Pre-/Posttests)= .27.
and the Managerial class at the .05 alpha level. The Typical class that included written tests as part of the grading system scored significantly higher than either of the other two classes that did not include written tests as part of the grading system.

The null hypothesis that the three grading systems would have no effects on the scores of the written examination was rejected.

Comparison of Grading Effects on Game Play Variables

As mentioned in the DIBA data, process measures of student achievement play an important role in teacher effectiveness research and research on teaching in physical education. Another important technique, besides the time-interval used in DIBA, in process measures is event recording (Darst et al., 1989). However, both the time-interval and event recording techniques have their disadvantages in certain circumstances. Nevertheless, a combination of the above two techniques overcomes their shortcomings while strengthening their advantages. Game play measuring skills are based on the above belief. In generating game play data, each student's motor attempts were recorded according to pre-defined categories within a defined amount of time. One of the most useful pieces of
Table 11. ANCOVA On Written Scores.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to Regression</td>
<td>1</td>
<td>204.02</td>
<td>15.18</td>
<td>.01</td>
</tr>
<tr>
<td>Post-Test</td>
<td>2</td>
<td>121.91</td>
<td>4.54</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>40</td>
<td>537.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>43</td>
<td>863.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12. Mean Comparison On Written Scores.

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean(Pretest)</th>
<th>SD(Pretest)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>14</td>
<td>6.13 (3.59)</td>
<td>3.60 (4.70)</td>
<td>S-T *</td>
</tr>
<tr>
<td>Managerial</td>
<td>13</td>
<td>5.73 (1.76)</td>
<td>3.58 (2.88)</td>
<td>M-T *</td>
</tr>
<tr>
<td>Typical</td>
<td>17</td>
<td>10.17 (4.51)</td>
<td>4.69 (4.74)</td>
<td>S-M</td>
</tr>
</tbody>
</table>

* = p<.05

r(Pre-/Posttests) = .49.
information provided by this kind of recording was the measure on rate of student behaviors due to the two dimensions about student motor attempts: time duration and number of events.

In this study, the duration of time was 11 minutes. Each student played 11 minutes, and was recorded five times. All motor attempts were initially categorized into three classes: serves, returns, and rallies. Furthermore, serves were divided into forehand and backhand which were classified as good serves, scored serves, and missed serves. Returns and rallies were divided into scored and missed (for detail, refer to Chapter 3). Results on game play are presented by the following categories: forehand good serves, forehand scored serves, forehand missed serves; backhand good serves, backhand scored serves, backhand missed serves; returns scored, returns missed; rallies scored, rallies missed. Then, these sub-categories were combined as backhand serves, forehand serves, total serves, returns, and rallies. Results on each of above sub-categories are presented in two separate tables. The first one is the ANOVA table and the second is the mean comparison table. The key to reading these tables is to read them as "student opportunity to respond in 11 minutes".

Table 13 presents ANOVA data on the forehand good serves for the three classes under three different grading
systems. Results of ANOVA showed that the number of forehand good serves significantly differed among the three classes, \( F(2, 192) = 12.18, p<.01 \). Within 11 minutes of playing time, a different class executed a different number of forehand good serves. Good serves are not necessarily related to missed returns of the opponent. Further mean comparisons (Table 14) showed that mean numbers of forehand good serves for the Skill, the Managerial, and the Typical classes, were 21.57, 16.03, and 19.63. Differences between the Skill and the Managerial, and the Typical and the Managerial classes were significant with alpha at .05. These data implied that students in the Skill (Mean = 21.57) and the Typical (Mean = 19.63) classes executed more forehand good serves in their 11-minutes of game play than those students in the Managerial class (Mean = 16.03).

All students in the three classes were informed that the criteria of success for game play had two dimensions: quantity and quality. Quantity refers to the number of games, rallies, serves, and returns. The more games, rallies, serves, and returns, the better. Quality refers to the numbers of games won, returns scored, and serves scored. The more games won, returns scored, and serves scored, the better. Students in the Managerial class were not graded on game play, only on attendance and participation. Meanwhile,
Table 13. ANOVA On Forehand Good Serve (FGS).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>1020.93</td>
<td>12.18</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>192</td>
<td>7959.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>194</td>
<td>8980.75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 14. Mean Comparison On Forehand Good Serve (FGS).

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>21.57</td>
<td>5.79</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>16.03</td>
<td>7.49</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>19.63</td>
<td>5.90</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

* = p<.05
game play accounted for 50% of the grade for the Skill class and 20% for the Typical class.

It should be noted, again, that there was a positive relationship between performance of forehand good serves and the connection of game play with student grades. The higher the percentage of game play in the explanation for grades, the more good forehand serves there were.

Next, results of ANOVA and mean comparisons on the forehand scored serves, forehand missed serves; backhand good serves, backhand scored serves, backhand missed serves; returns scored, returns missed; rallies scored, rallies missed are presented in Tables 15 to 32. Because information contained in Tables 15 to 32 can be read in a similar way as Table 13 and Table 14 on the forehand good serves, and because it is less meaningful to discuss information in each table separately (since information in an individual table is likely related to data in other tables), the information in these tables is presented in a combined approach.

Results of tables 15 to 32 fall into three groups according to the characteristics of significance on differences of the ANOVA and/or mean comparisons.

The first group is related to those tests which yield no significant conclusions. This group is presented in Tables 17, 18, 23, 24, 31, and 32. In other words,
information presented in these tables show that there were no significant differences among the three grading treatments in terms of forehand missed serves, backhand missed serves, and rallies missed.

The second group is related to those results which were significantly different among the three groups but the effects of treatments were not straightforward and were difficult to explain. These results related to forehand scored serves, backhand good serves, backhand scored serves, returns scored, and returns missed. These results are presented in Tables 15, 16, 19, 20, 21, 22, 25, 26, 29, and 30.

For example, Tables 15 and 16 show that the Skill class (Mean = 1.83) executed less forehand scored serves than the other two classes (Mean of the Managerial class = 4.01; Mean of the Typical class = 5.03), $F(2, 190) = 12.19$, $p<.01$. However, the fact that one scored more forehands may imply that one does not use backhands or misses more in using the backhand. Meanwhile, a class scoring high on forehand scored serves means that this class lost the same amount of points in returning forehand serves. This kind of information suggests that the Skill class needed more training on forehand serves and the Managerial and Typical classes needed to strengthen their capability to return forehand serves.
The final group contains Tables 13, 14, 27, and 28 (Tables 13 and 14 were presented and discussed prior to presentation of the first group as an example). Results of these tables can be explained in a straightforward way by the three treatment effects.

Next, results on total of forehand serves, backhand serves, serves, returns, and rallies totals are presented.

Tables 33 and 34 present results of ANOVA on the total number of returns which is comprised of returns missed and returns scored. In each 11-minute game, an average student in the Skill class returned significantly more times (Mean = 96.97) than an average student in both the Managerial (Mean = 80.53) and the Typical (mean = 76.48) classes, F(2, 190) = 9.80, p<.01. These results should be considered consistent with the effects of treatment although there is no difference between the Managerial and the Typical classes.

Tables 35 and 36 present results of ANOVA on the total number of rallies which is made up of rallies both scored and missed. Both the Skill class (Mean = 60.37) and the Typical class (Mean = 61.12) played significantly more (points) rallies than the Managerial class (mean = 53.38) in an average 11-minute game, F(2, 190) = 4.00, p<.01.
Table 15. ANOVA On Forehand Scored Serve (FSS).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>330.83</td>
<td>12.19</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>2577.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>2908.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 16. Mean Comparison On Forehand Scored Serve (FSS).

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>1.833</td>
<td>5.28</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>4.01</td>
<td>4.83</td>
<td>S-T *</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>5.03</td>
<td>3.32</td>
<td>M-T</td>
</tr>
</tbody>
</table>

* = p<.05
Table 17. ANOVA On Forehand Missed Serve (FMS).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>6.23</td>
<td>0.61</td>
<td>.55</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>976.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>982.52</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 18. Mean Comparison On Forehand Missed Serve (FMS).

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>0.93</td>
<td>2.15</td>
<td>S-M</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>1.28</td>
<td>1.76</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>1.35</td>
<td>2.78</td>
<td>M-T</td>
</tr>
</tbody>
</table>
Table 19. ANOVA On Backhand Good Serve (BGS).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>91.03</td>
<td>2.25</td>
<td>.11</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>3847.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>3938.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 20. Mean Comparison On Backhand Good Serve (BGS).

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>4.58</td>
<td>5.24</td>
<td>S-M</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>3.63</td>
<td>4.84</td>
<td>S-T *</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>2.88</td>
<td>3.20</td>
<td>M-T</td>
</tr>
</tbody>
</table>

* = p<.05
Table 21. ANOVA On Backhand Scored Serve (BSS).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>22.70</td>
<td>5.11</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>422.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>444.72</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 22. Mean Comparison On Backhand Scored Serve (BSS).

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>0.38</td>
<td>0.87</td>
<td>S-M</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>1.16</td>
<td>1.94</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>1.08</td>
<td>1.40</td>
<td>M-T</td>
</tr>
</tbody>
</table>

* = p<.05
Table 23. ANOVA On Backhand Missed Serve (BMS).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>0.76</td>
<td>0.43</td>
<td>.65</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>168.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>169.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 24. Mean Comparison On Backhand Missed Serve (BMS).

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>0.35</td>
<td>0.80</td>
<td>S-M</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>0.50</td>
<td>1.04</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>0.40</td>
<td>0.95</td>
<td>M-T</td>
</tr>
</tbody>
</table>
Table 25. ANOVA On Return Scored (RES).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>3571.69</td>
<td>6.34</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>53479.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>57051.49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 26. Mean Comparison On Return Scored (RES).

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>48.43</td>
<td>16.30</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>40.09</td>
<td>16.09</td>
<td>S-T *</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>38.4</td>
<td>17.88</td>
<td>M-T</td>
</tr>
</tbody>
</table>

* = p<.05
Table 27  ANOVA On Rallies Scored (RAS).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>648.66</td>
<td>3.28</td>
<td>.04</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>18775.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>19424.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 28. Mean Comparison On Rallies Scored (RAS).

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>30.12</td>
<td>7.95</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>26.51</td>
<td>10.99</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>30.54</td>
<td>10.42</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

* = p<.05
Table 29  ANOVA On Returns Missed (RES).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>3721.71</td>
<td>5.11</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>69140.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>72862.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 30  Mean Comparison On Returns Missed (RES)

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>48.53</td>
<td>17.02</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>40.44</td>
<td>17.84</td>
<td>S-T *</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>38.08</td>
<td>21.92</td>
<td>M-T</td>
</tr>
</tbody>
</table>

* = p<.05
Table 31  ANOVA On Rally Missed (RAM)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>557.70</td>
<td>2.89</td>
<td>.0582</td>
</tr>
<tr>
<td>Residual</td>
<td>190</td>
<td>18353.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>18910.70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 32  Mean Comparison On Rally Missed (RAM)

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>30.25</td>
<td>7.89</td>
<td>S-M</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>26.87</td>
<td>10.83</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>30.57</td>
<td>10.32</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

* = p<.05
Table 33. ANOVA On Total of Returns.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>14571.36</td>
<td>9.80</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>141265.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>155836.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 34 Mean Comparison On Total of Returns

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>96.97</td>
<td>24.98</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>80.53</td>
<td>29.47</td>
<td>S-T *</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>76.48</td>
<td>26.89</td>
<td>M-T</td>
</tr>
</tbody>
</table>

* = p<.05
### Table 35: ANOVA On Total of Rallies

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>2409.10</td>
<td>5.73</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>39934.24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TOTAL**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TOTAL</strong></td>
<td>192</td>
<td><strong>42343.34</strong></td>
</tr>
</tbody>
</table>

### Table 36: Mean Comparison On Total of Rallies

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>60.37</td>
<td>10.12</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>53.38</td>
<td>17.20</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>61.12</td>
<td>14.82</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

* = p<.05
### Table 37 ANOVA On Total of Forehand Serves

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>752.82</td>
<td>6.12</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>11679.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>192</td>
<td>12432.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 38 Mean Comparison On Total of Forehand Serves

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>24.33</td>
<td>6.29</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>21.32</td>
<td>10.03</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>26.01</td>
<td>6.37</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

* = p<.05
Table 39  ANOVA On Total of Backhand Serves.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>38.99</td>
<td>0.61</td>
<td>.54</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>6065.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>192</td>
<td>6104.95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 40  Mean Comparison On Total of Backhand Serves

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>5.32</td>
<td>5.77</td>
<td>S-M</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>5.29</td>
<td>6.54</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>4.35</td>
<td>4.40</td>
<td>M-T</td>
</tr>
</tbody>
</table>
### Table 41 ANOVA On Total of Serves

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>2</td>
<td>257.16</td>
<td>4.37</td>
<td>.01</td>
</tr>
<tr>
<td>Residual Error</td>
<td>190</td>
<td>11466.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>192</td>
<td>11994.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 42 Mean Comparison On Total of Serves

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>60</td>
<td>29.65</td>
<td>4.49</td>
<td>S-M *</td>
</tr>
<tr>
<td>Managerial</td>
<td>68</td>
<td>26.62</td>
<td>10.40</td>
<td>S-T</td>
</tr>
<tr>
<td>Typical</td>
<td>65</td>
<td>30.37</td>
<td>6.88</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

* = p<.05
In addition, the same trend was found related to the total of forehand serves (Tables 37 and 38), $F(2, 192) = 6.12, p<.01$, although there was no significant difference found in terms of the total number of backhand serves (Tables 39 and 40). With total number of forehand serves, both the Skill and the Typical classes scored significantly higher than the Managerial class.

After combining the total number of forehand serves and backhand serves (Tables 41 and 42) and analyzing these data, the Skill class (Mean = 29.65) and the Typical class (Mean = 30.37) served significantly more times than the Managerial class (Mean = 26.62), $F(2, 190) = 4.37, p<.01$.

The major treatment effects in game play are summed up in Table 43. In reading Table 43, (a) "h" indicates that the mean of this treatment was higher than at least one of the other two treatments, (b) "l" indicates the mean of this treatment was lower than at least one of the other two treatments, (c) "N" indicates that there was not any significant difference, (d) "E" indicates that the data on this variable showed the effects of the treatments, (e) "S" indicates that the data on this variable showed strong effects of the treatments.
Table 43

Major treatment effects on the variables in game play.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Skill</th>
<th>Managerial</th>
<th>Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forehand Good Serves</td>
<td>S</td>
<td>h</td>
<td>1</td>
</tr>
<tr>
<td>Forehand Scored Serves</td>
<td>N</td>
<td>l</td>
<td>h</td>
</tr>
<tr>
<td>Forehand Missed Serves</td>
<td>N</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Backhand Good Serve</td>
<td>N</td>
<td>h</td>
<td>1</td>
</tr>
<tr>
<td>Backhand Scored Serves</td>
<td>N</td>
<td>l</td>
<td>h</td>
</tr>
<tr>
<td>Backhand Missed Serves</td>
<td>N</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Return Scored</td>
<td>E</td>
<td>h</td>
<td>h</td>
</tr>
<tr>
<td>Rallies Scored</td>
<td>S</td>
<td>h</td>
<td>1</td>
</tr>
<tr>
<td>Returns Missed</td>
<td>N</td>
<td>h</td>
<td>1</td>
</tr>
<tr>
<td>Rallies Missed</td>
<td>N</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total of Returns</td>
<td>E</td>
<td>h</td>
<td>1</td>
</tr>
<tr>
<td>Total of Rallies</td>
<td>S</td>
<td>h</td>
<td>1</td>
</tr>
<tr>
<td>Total/forehand Serves</td>
<td>S</td>
<td>h</td>
<td>1</td>
</tr>
<tr>
<td>Total/backhand Serves</td>
<td>N</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total of serves</td>
<td>S</td>
<td>h</td>
<td>1</td>
</tr>
</tbody>
</table>
Results of ANOVA and mean comparison tests indicated that the following variables in game play showed Strong Effects of the Treatments; forehand good serves, rallies scored, total of rallies, total of forehand serves, and total of serves. Strong Effects of the Treatments refer to those significant differences that existed in the mean comparisons for at least two pair of classes and the differences were positively related to treatments. For example, results on forehand good serves showed strong effects of the treatments because the Managerial class (Mean = 16.03) had lower points than both the Skill class (Mean = 21.57) and the Typical class (Mean = 19.63) while only the students in the Managerial class were not graded on performance in game play (Tables 13 and 14).

The Effects of Treatments means that significant differences existed in the mean comparisons for at least one pair of classes and the differences were positively related to treatments. Variables which showed effects of treatments were returns scored and total of returns.

Because students in both the Skill and the Typical classes served more times and made more points than those students in the Managerial class and students in the Skill class had more returns than those students in the Managerial class, the null hypothesis that the three treatments have no effects on student performance in game play is rejected. In
other words, because there were significant differences in the three most important variables (serves, returns, and rallies), there is confidence to state that those students in the Skill and the Typical classes displayed more opportunities to play table tennis which was influenced directly by the demands of their grading systems.

Comparisons of Student Attitude Toward the Three Grading Systems

Affective outcomes are considered an important product of physical education (Dick & Carey, 1990; Mao & Zakrajsek, 1989). In addition, it is assumed that student attitudes toward different grading systems are meaningful in understanding student behaviors and promoting student achievement. Furthermore, it is also documented that different grading systems have different effects on student perceptions and attitudes toward those grading systems (Dunham, 1986; Harrison, 1983; Neilson & Jensen, 1972; Safrit, 1973).

In this study, a 15-item questionnaire was designed to measure student perceptions on and attitudes toward these three different grading systems. Results of student ratings on each item are presented in Table 44 (for detail, refer to Appendix G). Each student was asked to select one of five choices: (a) strongly agree, (b) agree, (c) neutral, (d) disagree, and (e) strongly disagree where "Strongly agree"
was converted into "1" point, "agree" = two points, and so on. Therefore, the lower the response score, the more likely the student agreed with that statement.

Results of mean comparisons on the first question "I liked this grading system." showed that students under the Typical grading system kept a slightly neutral position (Mean = 3.33, 3 = Neutral), the students under the Skill grading system tended to agree (Mean = 2.35), while students under the Managerial grading system strongly agreed (mean = 1.43). It seems that students liked the Managerial grading system most, followed by Skill grading and Typical grading systems.

A review of the rest of the questionnaire shows that student perceptions on the first question tended to be similar to the other 13 questions with the exception of question 7; there was no difference among the three classes related to question 7.

In order to avoid discussing similar results on these 13 questions, it can be stated that the results of a principal component analysis confirmed the aforementioned statement that student responses on all 15 items (except item seven) fell into two patterns; responses on question seven and responses on all the other 14 questions (Table 45). In other words, there were only two factors which
Table 44. Mean Comparisons of Student Perceptions on the Three Grading Systems.

<table>
<thead>
<tr>
<th>Questions</th>
<th>S/M</th>
<th>S/T</th>
<th>M/T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this grading system.</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>2. This grading system was explained clearly.</td>
<td>&gt;</td>
<td>-</td>
<td>&lt;</td>
</tr>
<tr>
<td>3. This grading system was fair.</td>
<td>&gt;</td>
<td>-</td>
<td>&lt;</td>
</tr>
<tr>
<td>4. Under this grading system, I learned skills better.</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>5. Under this grading system, I learned more about table tennis</td>
<td>-</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>6. This grading system reflected a true understanding of my capabilities.</td>
<td>-</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>7. This grading system differentiated among students.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8. This grading system was consistent with course objectives.</td>
<td>&gt;</td>
<td>-</td>
<td>&lt;</td>
</tr>
<tr>
<td>9. Under this grading system, it was easier for me to achieve the grade I want.</td>
<td>&gt;</td>
<td>-</td>
<td>&lt;</td>
</tr>
<tr>
<td>10. Under this grading system, I worked to achieve my desired grade.</td>
<td>&gt;</td>
<td>-</td>
<td>&lt;</td>
</tr>
<tr>
<td>11. This grading system provided guidance for learning.</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>12. I prefer that more physical education classes use this kind of grading system.</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>13. This grading system reflected course content learning expectations.</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>14. This grading system was fair and realistic.</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>15. I recommended this grading system for physical activity courses.</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
</tbody>
</table>

"<" or ">" = "smaller than" or "larger than".
"S/M" = "mean comparison between the Skill and the Managerial classes". "T" = Typical class.
"-" = "no significant difference"
explained more than 75% of the total variance in responses on the 15 questions. Put in another way, these two factors were able to represent more than 75% of the information contained in the responses related to the 15 questions. According to Table 45, factor 1 represented all questions except question 7 which assumed factor 2.

It seems reasonable to name factor 1 "student good feelings" on different grading systems because factor 1 had high loadings on all the 14 questions and factor 2 as "This grading system differentiated among students." or question 7.

By use of these two factors, information contained in Table 44 related to 15 questions could be concisely stated as: Students under the Managerial grading system had the strongest good feelings, students under the Skill grading system had strong good feelings, and students under the Typical grading system had the least good feelings about the way in which they were graded. It should be noted that there was no difference on factor 2 or Question 7 among the three classes.

The null hypothesis that the three grading systems had no effect on student attitudes toward the three grading systems was rejected.
Table 45
Principle Component Analysis of 15 Questions.

<table>
<thead>
<tr>
<th>Questions</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this grading system.</td>
<td>.8714</td>
<td>--</td>
</tr>
<tr>
<td>2. This grading system was explained clearly.</td>
<td>.7644</td>
<td>--</td>
</tr>
<tr>
<td>3. This grading system was fair.</td>
<td>.9030</td>
<td>--</td>
</tr>
<tr>
<td>4. Under this grading system, I learned skills better.</td>
<td>.8367</td>
<td>--</td>
</tr>
<tr>
<td>5. Under this grading system, I learned more about table tennis</td>
<td>.8618</td>
<td>--</td>
</tr>
<tr>
<td>6. This grading system reflected a true understanding of my capabilities.</td>
<td>.8041</td>
<td>--</td>
</tr>
<tr>
<td>7. This grading system differentiated among students.</td>
<td>--</td>
<td>.8608</td>
</tr>
<tr>
<td>8. This grading system was consistent with course objectives.</td>
<td>.7955</td>
<td>--</td>
</tr>
<tr>
<td>9. Under this grading system, it was easier for me to achieve the grade I want.</td>
<td>.8318</td>
<td>--</td>
</tr>
<tr>
<td>10. Under this grading system, I worked to achieve my desired grade.</td>
<td>.6653</td>
<td>--</td>
</tr>
<tr>
<td>11. This grading system provided guidance for learning.</td>
<td>.8345</td>
<td>--</td>
</tr>
<tr>
<td>12. I prefer that more physical education classes use this kind of grading system.</td>
<td>.9274</td>
<td>--</td>
</tr>
<tr>
<td>13. This grading system reflected course content learning expectations.</td>
<td>.9129</td>
<td>--</td>
</tr>
<tr>
<td>14. This grading system was fair and realistic.</td>
<td>.9244</td>
<td>--</td>
</tr>
<tr>
<td>15. I recommended this grading system for physical activity courses.</td>
<td>.9466</td>
<td>--</td>
</tr>
</tbody>
</table>

Final Communality estimates: Total = 11.29 out of 15
Major results of the data analysis about the effects of the three grading systems on student table tennis performance, cognitive learning, and grading attitude are summed up in Table 46. Two concepts are necessarily needed to read Table 46. Strong Effects of the Treatments means that significant differences existed in the mean comparisons for at least two pairs of classes in that the measure or variable and the differences were positively related to treatments. The Effects of Treatments means that the significant differences existed in the mean comparisons for at least one pair of classes and the differences were positively related to treatments.

In this experimental study, strong effects of treatments were found to be related with the majority of the measures. These measures contain student attendance rate, one-minute volley test, serve test, written test, some variables on game play, and some questions on student attitudes. In addition, effects of treatments were found to be related to some of the variables on game play and some aspects of student attitudes. No significant differences were related to student behaviors coded on DIBA, some variables on game play, and some items on student attitude questions. Based on these results, all of the null hypothesis stated in Chapter 3 are rejected.
Table 46

Summary Table on the Effects of the Three Grading Systems on Different Student Outcome and Process Measures.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Effects of Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Behaviors by DIBA</td>
<td>N</td>
</tr>
<tr>
<td>Student Attendance Rate</td>
<td>S</td>
</tr>
<tr>
<td>One-minute Volley Test</td>
<td>S</td>
</tr>
<tr>
<td>Serve Skill Test</td>
<td>S</td>
</tr>
<tr>
<td>Written Test</td>
<td>S</td>
</tr>
<tr>
<td>Variables on Game Play</td>
<td>S E N</td>
</tr>
<tr>
<td>Attitude Questions</td>
<td>S E N</td>
</tr>
</tbody>
</table>

S = strong effects of treatments
E = effects of treatments
N = no significant differences
CHAPTER V
SUMMARY AND DISCUSSION, CONCLUSION, AND RECOMMENDATION

Summary and Discussion

This research study tested and identified the effects of three grading systems on student outcome measures. The means to accomplish the above goal was a quasiexperiment in which the three grading systems were the independent variables and student outcome measures were the dependent variables. Three table tennis classes in a basic instruction program were selected as the experimental units. The study lasted the entire quarter in which all of the instructional variables except grading, were kept as equal as possible across the three classes. The three grading systems were based on theory and practice and a careful review of literature. The three grading systems employed in this experiment were:

1. Skill grading system under which students were graded totally on skill outcomes: (a) 25% from a one-minute volley test, (b) 25% from a serve test, and (c) 50% from scores on game play variables.
2. Managerial grading system under which students were graded totally on attendance and participation which can be simply explained as "attending class and being on task".

3. Typical grading system under which students were graded on four components: (a) attendance and participation (40%), (b) skills, one-minute volley and serve tests, (20%), (c) written tests (20%), and (d) game play (20%).

Student outcome measures were related to student behaviors defined by DIBA, attendance rates, one-minute volley and serve tests, written examination, game play variables, and an attitude questionnaire. Statistical results (with some explanation) of this experimental study on the effects of these three grading systems on the above student outcome measures were presented in the previous chapter. To further facilitate discussing and interpreting these statistical results, these findings will be viewed through three perspectives in this chapter: (a) positive effects of grading systems, (b) reversed effects of grading systems across attitude measures and other outcomes, and (c) non-significant measures on effects of grading systems.

Positive Effects of Grading Systems

The null hypotheses that there would be no grading system effects on student outcome measures related to attendance rates, two skill tests, written examination, game play variables, and attitude questionnaire were rejected.
In other words, it was found that there were grading effects on student outcome measures. Effects of these grading systems were found positively related to student outcome measures with the exception of grading effects on the measures of the attitude questionnaire which will be discussed later. Positive effects meant that students in the Skill grading class had significantly higher scores on the one-minute volley and serve skill tests, and the major game play variables (total of returns, rallies, and serves) than students under the Managerial grading system; and, they had significantly higher total returns in game play than students under the Typical class. Positive effects of grading showed that the Managerial class had significantly higher attendance rates than either of the other two classes. Finally, positive effects of grading systems indicated that the Typical class had significantly higher written test scores than the other two classes and had significantly higher scores on the one-minute volley test, serve test, total of rallies, and total of serves than the Managerial class; and, they also had a significantly higher attendance rate than the Skill class. In short, a causation between these student outcome measures and the grading systems was demonstrated.

Positive effects of grading systems on student outcome measures can be initially interpreted in light of either the
Dunkin & Biddle model (1974) or the popular process-product paradigm. There are three kinds of variables (presage variables, context variables, and process variables) related to student outcomes (product variables). Although Dunkin & Biddle do not specify grading systems in their variable categories, it seems plausible that a grading system could be considered as one of the process variables. In the process-product paradigm, two kinds of variables (process and product) and their causation were studied which were discussed in the Dunkin & Biddle model. It should be noted that the design of this study had its underpinnings in the Dunkin & Biddle model and the process-product paradigm because the three grading systems were taken as one kind of process variable with student measures reflecting the products. In teacher effectiveness research, many process variables were identified as consistently related to student achievement (Brophy & Good, 1986).

Furthermore, the positive effects of grading systems can be explained through the notion of "exchange of performance for grades" which is a key concept in the ecological model in both research on teaching and research on teaching in physical education. Exchange of performance for grades simply refers to the belief that in the problematic school setting, the purpose of students' performance is to trade performance for a good grade.
Therefore, academically, students are socialized into performing assigned tasks at some level commensurate with their expectations of grades as directed by the grading system.

In addition, the positive effects of grading systems was somewhat consistent with learning outcomes inasmuch as these students were motivated to learn by external factors. When grades did not include demonstrated motor and cognitive learning, these students were less motivated to excel.

One of these positive effects of grading systems was consistent with results of similar experimental studies in physical education, which tested the causation of motor performance and grading systems. Boyce's positive effects of grading systems were only related to motor skill measures (Boyce, 1990). In another study, no significant positive grading effects were identified (Mueller, et al, 1988) although the authors suggested that there was evidence about grading effects. Results of the grading systems identified in this study not only confirm the findings from Boyce's study and the beliefs of Mueller, et al., but they expand previous findings into other domains: attitudes, verbal information, intellectual skills, and cognitive strategies. Success on the written test in this study required using verbal information, intellectual skills, and cognitive strategies. It has been widely accepted that physical
education affects all learning domains; motor skill, attitudes, verbal information, intellectual skills, and cognitive strategies (Gagne, et al., 1988). These results on grading effects have important theoretical and practical meanings in that grading in four of the five domains showed that how much students learned was influenced by a particular grading system.

Reversed Effects of Grading Systems Across Attitude and the Other Outcome Measures

Student outcomes are the major criteria in terms of teaching effectiveness in research on teaching. Student outcomes in physical education are usually condensed into three domains; cognitive, affective, and psychomotor, although some authors include expended areas (Pangrazi & Darst, 1985). An optimal situation may be one in which the positive effects of the grading system are shown consistently in all domains of the three major categories (cognitive, affective, and motor skill). However, it is more likely that the situation is one in which the effects of different processes on the product are inconsistent. In teacher effectiveness studies, although consistent results are found in some studies in terms of student cognitive and affective outcomes, inconsistent findings within these two domains are often noted. Sometimes, the effects of the
different process (independent) variables on the product (dependent) variables are reversed.

In this study, reversed effects of grading systems were found across the attitude measures and some other outcome measures related to the two skill tests, written test, and variables on game play. Although students under the Managerial grading system generally had lower scores on both skill tests and the major variables of game play than students under the Skill and Typical grading systems and lower written test scores than students under the Typical grading system, they had more positive attitudes as measured by the questionnaire (Table 44). It seems contradictory that those students under the Managerial grading system responded to the above questions in a more favorable way although they scored lower on skill tests, written test, and on the variables of game play.

Furthermore, there were four categories of course objectives listed on the Course Outline, which were related to student outcomes in all of the five domains. Under the Managerial grading system, student learning was less consistent with course objectives, but they felt better about how they were evaluated.

Technically, according to the principal component analysis, all 15 questions could be economically represented by two factors or grouped into two groups, factor 1 and
factor 2. Factor 2 had only one question (question 7) in it and factor 1 had the remaining items. The 14 questions had similar characteristics about student feelings on grading and were collectively labelled "student good feelings" about grading systems. Therefore, students under the Managerial grading system had the strongest good feelings toward that grading system, but had the poorest learning scores.

Similar reversed effects of grading systems were shown in comparisons between the Typical and Skill grading systems across some of the student outcome measures. Students under the Typical grading system had higher attendance rates and written test scores than students under the Skill grading system, yet they showed less good feelings about their (Typical) grading system than their counterparts under the Skill grading system.

In short, the reversed effects of grading systems across attitude measures and other outcome measures means that students had less good feelings about those grading systems which had more positive effects on their learning other outcome measures. Although the effects of the three grading systems across attitudes and the other outcome measures were reversed in the aforementioned way and when viewed through the notion of an exchange of performance for grades, the effects of the three grading systems across attitude and the other outcome measures were consistent in
the following sense. In order to secure a good grade, students knew that higher demands from a grading system meant more work, or otherwise, they risk the possibility of obtaining a good grade. This notion was well explained by Doyle's (1979) concepts; ambiguity and risk. Ambiguity refers to the gap between the requirements of tasks and the information available to deal with those tasks. Risk means the possibility of securing a grade weighted against losing that grade. It is obvious that the three grading systems (the Skill, Managerial, and Typical) created different degrees of ambiguity and risk. It seems that the Typical grading system created more pressure on students than the other two grading systems because it covered more areas (skill, game play, attendance, participation, and written). The Managerial grading system was the least demanding by only requiring full attendance and participation to secure an "A" while full attendance and participation would only earn 40% of the grade under the Typical grading system and zero percent under the Skill grading system. However, in order to obtain an "A", students in the Skill and Typical classes needed to accomplish more tasks, especially students under the Typical grading system.

It was not too surprising that students had the least good feelings toward the Typical grading system and the most good feelings toward the Managerial grading system.
Even though the effect of grading systems on the attitude and the other product measures appeared in opposite directions, they were consistent with the notion of an exchange of performance for a grade.

In addition, it is likely that the reason students liked the Managerial grading system better is related to their experience with grading practices in schools. As shown in the review of literature on grading in physical education, there is a strong trend to grade students according to managerial aspects rather than learning outcomes. Growing up with this kind of expectation, students naturally expressed less good feelings about grading in physical education which placed higher demands on them.

**Non-Significant Measures on the Effects of Grading Systems**

The purpose of this quasiexperimental study was to test the effects of three grading systems on students' outcome measures. Some results of comparisons on these students' outcome measures showed that there were no significant differences across the three grading systems on: (a) student behaviors as defined by DIBA (student motor engaged, student cognitive engaged, student response preparing, student getting equipment or relocating, student waiting, and student off-task), (b) three variables on game play
(forehand missed serves, backhand missed serves, and total of backhand serves), and (c) item seven on the questionnaire; "This grading system differentiated among students".

Two kinds of measures (process and product measures) were employed in this study. Both the process and product measures provided good data in the sense that each measure (either a process measure or a product measure) constructed one aspect of the triangulation in data collection. However, as mentioned in Chapter IV when the data derived from DIBA were presented, both the process measures and product measures had their advantages and disadvantages. In this study, the process measures were related to use of DIBA and game play variables while those measures on skill tests, written test, attendance rates, and attitudes fell into the product measures. It was found that significant differences were related to both the product and process measures. However, the big picture is that (a) within the process measures, data on DIBA provided non-significant results while the game play measures identified important significant differences in major variables, (b) both the significant and non-significant results were related to product measures. Major data across the three classes on the skill tests, written test, attendance rates, and factor 1 on attitude were significantly different while some
variables and some comparisons among the skill tests, written test, and attendance rates were not significantly different.

Within the two process measures; DIBA and game play, the results were opposite in terms of statistical significance. The majority of the game play variables identified significant differences across the three grading systems, but data from DIBA showed that there were no significant differences across the three classes. The reason provided in Chapter IV was that DIBA was limited by time interval recording technique which only measured one dimension (time duration) while the game play variables measured two dimensions (duration and the number of motor attempts).

Limitations of time interval recording systematic observation instruments are widely noted (Alexander, 1983). According to interval recording techniques, the total amount of student time is divided into small intervals such as five-second or three-second intervals. Each of these intervals is classified into exclusively different categories of the system according to the major characteristics of each interval. This method does not tell the "quality of substance" in each interval such as the change rate or change speed. In other words, the interval technique measures duration; but duration does not measure
behavior rate, speed, events, latency, acceleration, and inter-response time.

Within this study, the teacher's use of time was well planned in order to keep it consistent across the three classes. The well planned time use by the teacher also indirectly determined the time use of students. For example, when 10 minutes were assigned to teacher lecturing, the same 10 minutes were picked up in student "cognitive engagement" because it was impossible for students to behave in other categories of DIBA. Therefore, a two dimensional technique such as game play methods (or techniques of measuring opportunity to respond) is recommended rather than a time interval technique, especially, in an experimental situation like this one in which the use of time was directly or indirectly pre-determined.

However, it should be noted, all occurrences of game play variables, temporarily, fell into one of the DIBA categories; student motor engagement. If there was any difference in the other categories such as student off-task and student cognitive engagement, the game play variable would be unable to unidentifiable. This is an example of the disadvantages that were related to a measuring technique on opportunity to respond such as these game play measures.

Non-significant results related to the game play variables (forehand missed serves, backhand missed serves,
and total of backhand serves) seems to simply imply that there were no differences across the three classes in terms of these variables. For backhand missed serves and the total of backhand serves, it was apparent that students in all three classes preferred using forehand serves. Very few backhand serves were used. The small sample size related to backhand serves was a major contributor for non-significance.

That there was no significant difference in the variable "forehand missed serves" indicated an important concern or difficulty in this study in the sense that it was difficult to interpret this variable. For example, if there were significant differences in the forehand missed serves across the three grading systems, what do the difference mean? Does it mean that students were under more stress, more tension because of the effects of the grading systems, or simply because they were less skillful? This similar concern or difficulty was related to most of the game play variables. One possible solution to this difficulty in interpreting game play variables is to provide a setting on which all variables could be explained. For example, all game play variables could be comparatively interpreted if students in all three classes were organized in the same tournament and played against each other. In this study, understanding the meanings of these game play variables was
restricted within the concept of "opportunity to respond in 11 minutes" because data on these game play variables were collected within each of the three classes and, additionally each class was a unit of analysis. Three other aspects need comment; contribution to ETU studies, cost of measurement, and difficulty in measuring motor skill learning.

The Contributions to Experimental Studies in Physical Education

Methodologically, experimental studies are considered the last phase of a complete and sound research program within the notion of a descriptive-correlational-experimental loop in the teacher effectiveness research. Experimental studies are usually conducted on the knowledge and questions derived from previous descriptive and correlational studies, therefore an experimental research study usually requires a better research base and sounder methodology (Kennedy & Bush, 1985; Shavelson, 1981). Most experimental research studies in physical education usually refer to experimental teaching unit studies (ETU studies).

Both the findings and methods related to this study have applicable meanings to ETU research. Advantages and disadvantages of an ETU study exist in its short duration of treatment, low cost, and small scale nature. So far, the contributions of ETU studies are limited within replications of some results done in other research paradigms. In
addition, the major findings derived from ETU studies are that there are no significant effects of the independent variables on the dependent variables (Graham, et al, 1983; Metzler, 1983; Pieron and Graham, 1984). However, the duration of treatments in this study was longer with positive causation identified between the grading systems and most measures on student outcomes.

**Cost of Measurement**

Since physical education affects student outcomes in all domains (motor skill, intellectual skill, cognitive skill, verbal information, and attitude), it seems desirable to relate all domains to the components of a grading system in a physical education course. In addition, students need to know that data on all components related to the grading system will be collected and that they will be held accountable. In other words, in order to implement a grading system, administration of pretests, posttests, process measures, and product measures to obtain data in assigning grades becomes a necessity. However, administration of tests is costly in terms of time. Regardless of good organization, several days may be required to conduct skill pre- and posttests, a written test, and an attitude survey. It is obvious that time is a practical obstacle in implementing a grading system which has components related to all domains.
Difficulty of Measuring learning Table Tennis Skills

Although some do not regard learning motor skills as the most important objective in physical education (Mao & Zakrajsek, 1988), learning table tennis skills was defined as one of the major objectives for all three classes. The content covered in these three classes centered on learning skills and applying them in game situations.

The difficulty in measuring table tennis skills stemmed from trying to measure the proficiency of each skill and the elements within each skill in a systematic and coherent way. Additionally, there was only one skill test (one-minute volley) available. The second test (serve), and game play measures were developed.

The degree to which the two skill tests matched the content covered seems problematic. First, the one-minute volley tested student volley proficiency in a way that was never taught and never practiced because they were required to volley against a perpendicular board. Second, with the serving test, students might get high scores at the cost of other elements (skill form, speed, spin, and trajectory) of a good serve. Therefore, developing sound skill tests is difficult, maybe impossible. Furthermore, from the point of view of the skills covered, the meaning of the measures on game play variables seems ambiguous. For example, the linkages between number of returns and learning a forehand
attack skill in terms of its five elements (skill form, speed, spin, trajectory, and landing) are uncertain. Although the number of returns reflects the opportunities to respond for a student, this measure is not directly related to skill learning in the way it was taught.

Conclusions

The purpose of this study was to identify the effects of three grading systems. The three grading systems were; (a) the Skill grading system in which students were graded 100% on skill tests, (b) the Managerial grading system in which students were graded 100% on attendance and appropriate participation, and (c) the Typical grading system in which students were graded on skill tests (40%), written test (20%), and attendance and appropriate participation (40%). Positive effects of these three grading systems were displayed on most student process and outcome measures (student attendance rates, skill tests, written test, and game play variables), however, a reversed effect of these grading systems was related to student attitude. These three grading systems had no significant effects on student behaviors as defined by a three-second interval instrument and some game play variables.

Conclusions about the effects of the three grading systems
which were based on three table tennis classes and the relevant findings are listed as follows.

1. Under the Skill grading system, students scored significantly higher than students under the Managerial grading system on the one-minute volley test.

2. Under the Skill grading system, students scored significantly higher than students under the Managerial grading system on the serve test.

3. Under the Skill grading system, students displayed significantly more opportunities in game play than students under the Managerial grading system.

4. Under the Skill grading system, students displayed significantly more opportunities in terms of returns in game play than students under the Typical grading system.

5. Under the Typical grading system, students had significantly higher scores on two skill tests than students under the Managerial grading system.

6. Under the Typical grading system, students displayed significantly more opportunities in game play than students under the Managerial grading system.

7. Under the Typical grading system, students had significantly higher scores on the written test than students under both the Managerial and the Skill grading systems.
8. Under the Managerial grading system, students had significantly higher attendance rates than the other two grading systems.

9. Under the Managerial grading system, students had significantly lower scores on two skill tests than students under both the Skill and the Typical grading system.

10. Under the Managerial grading system, students displayed significantly fewer opportunities in game play than students under both the Skill and the Typical grading systems.

11. Under the Managerial grading system, students had significantly lower scores on the written test than students under the Typical grading system.

12. Under both the Typical and the Skill grading systems, students had no significant differences in terms of one-minute volley test and serve test.

13. The effects of the three grading systems are reversed significantly on student attitude measures.

14. Under the Managerial grading system, student achievement on the skill tests, written test, and the game play measures are the poorest while they liked this Managerial grading system most.

15. Students had significantly more good feelings toward the Skill grading system than on the Typical grading system.
16. The more components a grading system contains, the more areas in which student learning takes place but the students like it less. In other words, the higher the demands of the grading system the more students are forced to learn, but they have less good feelings toward it.

Recommendations for Future Research

Grading is one of the most important issues or independent variables in school life and has important meaning to both teachers and students. We still have much to study about grading in terms of student learning in physical education. Suggestions and recommendations derived from this study for future research and practice are presented next.

Recommendations for Improving This Study

1. Separate the investigator and the teacher because only a separation of the teacher and the investigator can completely eliminate reactivity of the investigator in teaching.

2. Use a live coder or a camera operator to focus on the behaviors of both the teacher and the target student.

3. Collect data on whether and how much, under each of the three grading systems, students played table tennis outside classes to see the effects of these three grading systems on student behaviors outside classes.
Recommendations for Replicating This Study

1. Replicate this study in different activity areas in college basic instruction programs and in secondary schools.

2. Replicate this study using a pass and fail grading system against a letter grading system.

3. Replicate this study using gender as a second independent variable.

4. Use other systematic observation instruments to collect data on the dependent variables. For example, dependant variables may be entered through an instrument related to task structures and task congruence.

5. Study the effects of grading on teaching behaviors. It seems equally important to test the effects of grading systems on teacher behaviors because a grading system is equally important to both teachers and students.

Recommendations for Grading Activity Classes in Physical Education

1. Initially, a sound grading system should cover all outcome domains because it is likely that the components of a grading system strengthen students' learning. However, in order to implement a multiple-component grading system, such as the Typical grading system, it seems unrealistic in a non-research setting because of the large amount of time needed. Therefore, developing a grading system with major components defined by course objectives that can easily be
measured in a small amount of time seems most practical.

2. Determining the weights of each component for a grading system favoring motor skill development if skill achievement is the major objective of physical education. It is difficult to measure motor skill learning in the same way skills are taught because of the lack of appropriate skill tests.

3. The uncertainty about which grading system to use will continue as long as there is uncertainty about what the objectives and goals of physical education are.

4. One grading system should be adapted by the institution so that all students in all physical education classes are graded by the same criterion.

Concluding Statements

Grading systems which were communicated to students prior to instruction had influence on all aspects of students' learning (cognitive, affective, and motor skill). Use of a specific grading system in university basic instruction programs will enhance student learning in a specific way.
REFERENCES


learning time-physical education 1982 revision coding
manual. Columbus, OH: The Ohio State University.


Soar, R., and Soar, R.M. (1979). Educational climate and
management. In P. Peterson and H. Walber (Eds.),
Research on Teaching: Concepts, Findings, and

of Health, Physical Education, and Recreation, 38(5), 35-
36.

Son, C.T. (1989). Descriptive analysis of task congruence in
Korean middle school physical education classes.
Unpublished Doctoral dissertation, the Ohio State
University.

learning. Journal of Teaching in Physical Education. 8
(3). 213-222.

Classroom Observation Evaluation 1972-1973 (SRI Protect

selected resources from ERIC. JOPERD, 61(3). pp.26-
27.

Terry, M.W., & Safrit, M.J., (1990). Measurement and
evaluation in professional physical education-a view
from the measurement specialists. JOPERD, 61(3).
pp.29-31.

Tinning, R.I. (1983). A task theory of student teaching:
development and provisional testing. Unpublished
doctoral dissertation, the Ohio State University.

Tousignant, M. (1982). Analysis of the task structures in
secondary physical education classes. Unpublished
doctoral dissertation, The Ohio State University.

Trimble, R.T., & Hensley, L.D. (1984). The general
instruction program in physical education at four-year
pp.92-89.


APPENDICES
APPENDIX A

APPROVAL FORM FOR INVolVEMENT OF HUMAN SUBJECTS
BEHAVIORAL AND SOCIAL SCIENCES
HUMAN SUBJECTS REVIEW COMMITTEE (HSRC)
THE OHIO STATE UNIVERSITY

Date April 19, 1991

RESEARCH PROTOCOL:

91B0054 AN EXPERIMENTAL DESIGN STUDY ON THE EFFECTS OF THREE GRADING SYSTEMS ON STUDENT OUTCOMES IN A BASIC INSTRUCTIONAL PROGRAM,
Dorothy M. Zakrajsek, Youxiang Mao, Health, Physical Education and Recreation

presented for review by the Behavioral and Social Sciences Review Committee to ensure proper protection of the rights and welfare of the individuals involved with consideration of the methods used to obtain informed consent and the justification of risks in terms of potential benefits to be gained.

The Committee action was:

_____ APPROVED  _____ DEFERRED*

_____ APPROVED WITH CONDITIONS*  _____ DISAPPROVED

X NO REVIEW NECESSARY

*CONDITIONS/COMMENTS:

It was determined by the Committee that the protocol was EXEMPT FROM REVIEW.

COMMENT: Exempting an activity from review does not absolve the investigator(s) of the activity from ensuring that the welfare of subjects in the activity is protected and that methods used, and information provided, to gain subject consent are appropriate to the activity.

Date: April 19, 1991  
(Chairperson)
APPENDIX B

COURSE OUTLINES, UNIT TEACHING PLAN,
FIRST WEEK TEACHING PLAN

164
(For the class under the Typical Grading System)

OSU
TABLE TENNIS
SCHOOL OF HPER
COURSE # 185.01
SPRING 1991
CALL # 13778-9/13779-4

INSTRUCTOR: Youxiang Mao
OFFICE: To be arranged
MAILBOX: Pomerene Hall/Larkins
OFFICE HOURS: To be arranged
EMERGENCY MESSAGE PHONE: 292-2504
MEETING PLACE: Larkings 119
TIME: 11:00 12:00 T/R

COURSE OBJECTIVES:

By the end of this quarter, the student will be able to:

1. Execute basic serving and returning skills of playing table tennis. These basic serving and returning skills: topspinning serve, underspinning serve, sidespinning serve, attack and/or push-block, and short chop for both forehand and backhand.
2. Demonstrate understanding about basic rules and knowledge on skills/strategies of playing table tennis by solving written questions. This content will be contained in OFFICIAL RULES OF TABLE TENNIS (Long's Book Store) and Handouts (available at Kinko's).

And, by the end of this quarter, the student will have had the opportunity to:

3. Apply skills, knowledge on skills, strategies, and rules in game situations by participating in a tournament which consists of six to eight matches.

4. Participate in the course as defined by the evaluation standards.

EVALUATION:

1. Grading system:
   
   A 90%-100%  B 80%-89%  C 70%-79%  D 60%-69%
   
   E Below 59%

2. Grading based on:

   1. 40% active and appropriate attendance/participation.
   Attendance/Participation consists of the following components: Physically present, dressed appropriately, punctual, remain until class is over, on task.

   2. 40% skill test

   4. 20% written test
3. Absences:

More than six absences = E

Each unexcused absence cost 30%/6 = 6%

Each excused absence = 3%

TEXT BOOK

Official Rules of Table Tennis. (Available at Long's Book Store, approximately $ 0.50).

Handouts (available at Kinko's).

SKILL TESTS

I  Serve (10%).

Each student will be required to select any one serving skill from both forehand and backhand and demonstrate the two serving skills by serving the balls so that each ball lands on any arranged quarter of the opponents side of the table surface with certain speed, spin and low path. Each student serves 10 times for each skill.

Each good landing is worth of 0.5%. For example, 10 good landings out of 10 trials = 5% and 1 good landing out of 10 trials = 0.5%.

II Skills demonstrated in playing game (20%).

In this part of the skill test, the results of each return, each forehand serve, and backhand serve of your majority of game play will be recorded. Results of the forehand serves and backhand serves will be categorized as Good Serves, Serves Scored, and Serves missed. The number of returns in each volley and the number of volleys in each
game will be categorized as Returns Scored and Returns Missed.

You will be graded in terms of the quantity and quality of your game play. Quantity refers to the numbers of games, volleys, serves, and returns. The more games, volleys, serves, and returns you do, the more practice you have, the higher your grade should be.

Quality refers to the numbers of games won, returns scored, and serves scored, etc. Numbers of games won, returns scored, and serves scored will be positively related with your grade while the numbers of games lost, returns missed, and serves missed will be negatively related with your grade.

You will be asked to keep results of your game play for each other.

You also will be grouped into different groups in game play according to your skill levels. You will expect that players with similar skill levels play against one another.

III. One-minute volley test (10%).

For this test, a board is placed perpendicular to the playing surface in the middle of the table (between the two sides), to serve as a backboard. You are given one minute to volley a ball against the backboard as many times as you can.

Testing procedure is as follows: at the signal "Go", you drop a ball to the table and rally it against the
perpendicular board as many times as possible in one minute. Any number of bounces on the playing surface is permitted. If you lose control of the ball, another may be taken from the surface of the table and dropped to the playing surface, and played. Hits on the perpendicular surface do not count if the ball is volleyed (that is, the ball must bounce at least once), you put your free hand on the table during or immediately preceding a hit, or the ball strikes the perpendicular surface on the net. The test score is the number of volleys you did. You will be assigned a grade according to the norm-curve established before.

WRITTEN TEST (20%).

This part will be based on the OFFICIAL RULES OF TABLE TENNIS, handouts, and what has been taught in the class dealing with understanding of basic rules and knowledge on skills/strategies of playing table tennis.
(For the class under the Managerial Grading System)

Note: COURSE OBJECTIVES, TEXT BOOK, SKILL TESTS, AND WRITTEN TEST are similar to those listed for the Typical class.

EVALUATION:

1. Grading system:
   A 90%-100%  B 80%-89%  C 70%-79%  D 60%-69%  E Below 59%

2. Grading fully based on: Active and appropriate attendance/participation. Attendance/Participation consists of the following components: physically present, dressed appropriately, punctual, remain until class is over, on task, and others.

3. Absences:
   More than six absences = E
   Each unexcused absence cost 100%/20 = 5%
   Each excused absences=3%

   Under SKILL TESTS, students were informed: In this class, although you will be graded fully on appropriate attendance and participation, you will be asked to take skill tests and written tests to provide information on whether you have met the course objectives.
COURSE OBJECTIVES, TEXT BOOK, SKILL TESTS, AND WRITTEN TEST are similar to those listed for the Typical class.

EVALUATION:

1. Grading system:
   - A 90%-100%
   - B 80%-89%
   - C 70%-79%
   - D 60%-69%
   - E Below 59%

2. Grading based on:
   1. Serve test (25%).
   2. One-minute volley test (25%).
   3. Skills demonstrated in playing game (50%).

3. Absences:
   - More than six absences = E (School policy).
TIME SCHEDULE

This TIME SCHEDULE provides guidance for each lesson in all three Table Tennis classes.

1. Objectives: The teacher will be able to identify levels of student entry behaviors. The student will be able to demonstrate an understanding of the research study involved in these three Table Tennis classes.

   Tasks of instruction:
   1) Brief students about the research study;
   2) Give volley test. Students work in groups of three. One takes test. The second one retrieves ball and gives time. The last one keeps scores.
   3) Give serving test. Format as above.
   4) Give written test.
   5) Give questionnaires on affective outcome.

2. Continue the tasks on Day 1.

3. Objectives: The student will be able to generate the outline of this course and both the teacher and students will have the opportunity to introduce themselves.

   Tasks of instruction:
   1) Teacher and each student introduce herself/himself verbally.
   2) Introduce course syllabus for each class.
   3) If there is any time left after the above two tasks are completed, students introduce themselves while enjoying playing table tennis.

Objectives: The student will be able to:

1) execute the appropriate way of gripping a racket vertically and horizontally.
2) state the factors related to a good service (RULE pp.10-13).
3) execute forehand topspinning serve.

Tasks of instruction:
1) skill form of vertical and horizontal grips. Advantages/disadvantages and historic factors about these two grips.
2) Explain and demonstrate factors related to a good service.
3) Explain and demonstrate skill form, spin, speed, landing, and trajectory of topspinning serve.

5. Objectives: The student will be able to:

1) execute forehand sidespinning and underspinning serve.
2) state the factors related with skill forms, spins, speeds, landings, and trajectories of both forehand sidespinning and underspinning serves.

Tasks of instruction:
1) Sidespinning and underspinning serves for forehand.

6. Objectives: The student will be able to:

1) execute backhand topspinning, underspinning, and sidespinning serves.
2) state the factors related with skill forms, spins, speeds, landings, and trajectories of these above backhand serves.

Tasks of instruction:
1) the three backhand serves.
2) demonstrate and explain the five factors (skill form, speed, etc) related to the three backhand serves.


Tasks of instruction: explain and illustrate the above rules.

8. Objectives: The student will be able to execute forehand attack skill and backhand attack or push-blocking skills.

Tasks of instruction: explain and demonstrate the above skills.

9. Objectives: the student will be able to:
1) execute correct skill forms for short chop in both forehand and backhand.
2) state the five factors related to chopping skills.
Tasks of instruction:

1) explain and demonstrate short chops for forehand and backhand.

2) explain and demonstrate the five factors related to short chops.

10. Objectives: the student will be able to reach those objectives listed from Day 3 to Day 9 if h/she missed before.

Task of instruction: review all the skills covered in previous classes.

11. Objectives: the student will be able to

1) state the strategies in playing table tennis.

2) be prepared for a seven-days tournament.

Tasks of instruction:

1) explain and demonstrate strategies in playing table tennis. These strategies contain: a. identify the weakness/strength of your opponents in terms of forehand/backhand return and serve. b. manipulate the five factors of your skills.

2) arrange tournament.

12-17. Tournament days.

The purpose of this tournament is to provide an opportunity for students to learn and practice skills/strategies learned before in real game situation. In addition, this tournament provides real situation for
students to apply their knowledge on rules, skills, and strategies.

18-19. Posttest Day. The purpose of posttest is to collect data on student achievement. The procedure for post-test is similar to pre-test.

Note: Since May 27 is Memorial Day (No classes), there will be only 19 meetings for classes meeting on Monday and Friday. Also, there is no Table Tennis on May 28.
FIRST WEEK (Day 1 & Day 2) TEACHING PLAN

A. Teaching objectives: 1) identify the levels of student entry behaviors skill levels; 2) brief the student about the research project involved in all three Table Tennis classes.

B. Tasks of instruction:

1) Brief students about the research study.

2) Give volley test. Students works in group of three. One takes test. The second one retrieves ball and gives time. The last one keeps scores.

3) Give serve test. Format as above.

4) Give written test.

5) Give questionnaire on affective outcome.

C. Information about the research project involved in these three Table Tennis classes: All the three Table Tennis classes this Quater, Spring 1991, are involved in a research study. The purpose of this study is to identify some relationship between instructional variables and student achievement such as serving skill development in playing table tennis. It is expected that the result of this study will be useful for those students in future Table Tennis classes and other similar physical education classes. Your enrollments in these classes are helpful and appreciated. Besides the aforementioned information, the
instructor, who is a co-researcher, is not allowed to provide any further information about this research to any student.

You may choose to drop this class at any time as specified by the University Dropping Policy.

D. Organization and grouping: After the information about the research project has been addressed, the skill tests will be explained and demonstrated. Next, students will be organized in group of three. Volley test will be conducted from the beginning of the test time since there is only one table available for this volley test. The written test and questionnaire should be completed before the serving test starts.

F. The second meeting of the classes will continue the above tests and tasks. Those who have completed all above tests and tasks on the first day do not have to show up on the second day.

Make sure you will be here for the third day class!
APPENDIX C

DIRECT INSTRUCTION BEHAVIOR ANALYSIS

(DIBA) CODING MANUAL
Direct instruction is a model of teaching that is conceptually and operationally based on a combination of behaviors that were synthesized from process/product research (Rosenshine & Furst, 1971). Fielding, Kaneenui, and Gersten (1983) refer to direct instruction as a tightly structured approach to teaching in which the learner is shown explicitly how to perform skills, has ample practice time to master skills, receives frequent feedback on learning progress, and experiences success at each learning phase. Rosenshine (1979) says that direct instruction includes clear learning goals, sufficient time for instruction and practice, monitored student performance, low-level cognitive questions, immediate and specific feedback, extensive content coverage, and learning goals and materials appropriate for the student's ability. Proponents of direct instruction claim that learning achievement is increased through active teacher participation and on-task student behavior.

Direct Instruction Behavior Analysis (DIBA) features those behaviors identified by direct instruction researchers...
(Good & Grouws, 1977; Rosenshine, 1976, 1979; Rosenshine & Furst, 1971) and interpreted for physical education by Graham and Heimerer (1981). DIBA was designed to collect data on teacher and student behaviors that can be used in analyzing teaching performance with the direct instruction teaching model.

CATEGORY DESCRIPTIONS

DIBA comprises 15 categories: 8 categories plus 3 subscripts to further define teacher behavior and 4 categories to describe student behavior.

Teacher Behavior
Teaching Informing (I)
Teacher tells, explains, demonstrates, reviews, or summarizes.

Teacher Observing (O)
Teacher silently observes, watches, or monitors student performance.

Teacher Structuring (S)
Teacher stresses objectives and important points, directs performance, or signals transitions.

Teacher Praise/Encouragement (P)
Teacher praises, commends, accepts, or encourages student performance or attempts.

Teacher Feedback (F)
Teacher gives feedback that is immediate, specific, task relevant (can include correct or incorrect responses).
Teacher Controlling (C)
Teacher uses disciplinary comments or actions to criticize or to justify authority.

None of the Above
Teaching behaviors are not related to the instructional process (teaching and learning).

Enthusiasm (E)
A subscript E is used with the teacher behaviors if they are emitted in a way that interest and excitement are apparent through tone of voice, gestures, or facial expressions.

Lack Clarity (C)
A subscript C is used with teacher behaviors (informing, structuring, questioning, or feedback) to denote unclear interactions. Subscript C should be based on student reaction (e.g., puzzled looks, questions, and confusion) in following directions.

Modeling (M)
A subscript M is used with teaching behavior (informing, structuring, questioning, and feedback) to denote those times when the teacher is showing or demonstrating nonverbally how to perform or how not to perform a skill.

Student Behavior
Motor Engaged (M)
Student is actively engaged in an appropriate motor task/activity.
Cognitive Engaged (C)
Student listens to or reads about subject matter and gains information.

Response Preparing (R)
Student gets ready to respond to a learning task (R); gets equipment or relocates (Rr); waits for a turn (Rw).

Off-Task (O)
Student is not engaged in an appropriate motor or cognitive task. Examples: Student is getting a drink of water or taking a break.

RECORDING PROCEDURES
Three-second interval observe and record procedures are used. Acceptable levels of interobserver reliability can be established following five or six hours of training.

Agreements
________________________________________ x 100
Agreements + disagreements

The following recording procedures govern the data collection.
1. Start recording when the teacher officially begins the class and stop when the teacher determines the close of class.

2. Record behaviors in 3-second intervals. Observe the teacher for 3 seconds and during the next 3 seconds record the behavior. Then observe the target student for 3 seconds
and during the next 3 seconds record the behavior. The
sequence is as follows:
* Observe teacher 3 seconds
* Record teacher behavior 3 seconds
* Observe target student 3 seconds
* Record student behavior 3 seconds
Ten behaviors, 5 teacher and 5 student, are recorded per
minute.
3. Select a target student at random. Record all student
behaviors using only the target student.
4. If you use more than one target student, only observe one
at a time.
5. Use a split-column format for recordings and enter them
horizontally (see Coding Form).
6. Record behavior changes if observed during the 3-second
internal and adjust by slowing the subsequent recording
pace.
7. Use subscripts depending on the information sought.
8. Under the Comments space, make notes on student/teacher
behaviors which you believe are unique.
SUMMARIZING AND INTERPRETNG THE DATA
Tally the number of behaviors for each teacher and student
category. Figure the percentage of each category by dividing
the category tallies by the number of total tallies (use the
DIBA summary form).
The percentages of teaching behaviors can be analyzed according to the direct instruction teaching model. The subscripts allow for a closer examination of those behaviors that are more elusive in the overall generalization of limited teacher behaviors. The student behavior percentages allow for an analysis of on- and off-task behavior and how much time the students are engaged in learning. Information for DIBA can be used to compare instructional effectiveness as defined by direct instruction.
# DIBA Summary Form

## Teacher behaviors

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Number of tallies</th>
<th>Percent*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informing (I)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observing (O)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structuring (S)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questioning (Q)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Praise/encourage (P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback (F)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controlling (C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None of the above (N)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Enthusiasm** (E) \[\frac{I+S+Q+P+E}{I+S+Q+P} \times 100\]

**Clarity** (C) \[\frac{I+S+Q+F}{I+S+Q} \times 100\]

**Modeling** \[\frac{I+S+Q+F}{I+S+Q} \times 100\]

## Student behaviors

<table>
<thead>
<tr>
<th>Behavior</th>
<th># of tallies</th>
<th>Percent*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor engaged (M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive engaged (E)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response preparing (R)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting equipment or relocating (Rt)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waiting for a turn (Rw)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off-task (O)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Divide the number of tallies per behavior by the total number of tallies.

** Divide the number of enthusiasm tallies by the total number of tallies for informing, structuring, questioning, praise, and feedback.

*** Divide the number of clarity tallies by the total number of tallies for informing, structuring, questioning, and feedback.

**** Divide the number of modeling tallies by the total number of tallies for informing, structuring, questioning, and feedback.
Direct Instruction Behavior Analysis (DIBA)
Coding Sheet

T S T S T S T S T S T S T S T S T S

Comments

Teacher behaviors

I Informing  F Feedback
O Observing  C Controlling
S Structuring N None of the above
Q Questioning (E) Enthusiasm
P Praise/encourage (C) Clarity
                (M) Modeling

Student behaviors

M Motor engaged  C Cognitive engaged
R Response preparing
Rt Getting equipment
Rw Waiting for a turn
O Off-task
APPENDIX D

WRITTEN TEST
I. True or false: (One correct choice=one point; One incorrect choice=negative one point; Make no choice = zero point.)

1. ____ In general, a topspinning ball moves faster than an underspinning ball.

2. ___ Tom plays Mary. A game is started by Tom's serve. Therefore, we know that it is Tom's turn to serve after the scores reaches 20 to 21.

3. ____ It is necessary to do the following when starting to serve a ball: the ball rests on the palm of the free hand which is open and flat.

4. ____ In any situation, a game shall be won by the player first scoring 21 points.

5. ____ We see more underspinning balls than topspinning balls in this class.

6. ____ A game might consist of the best of three matches or the best of five matches.

7. ____ In order to serve a short ball, generally speaking, a underspinning ball is better than a topspinning one.

8. ____ In a formal tournament, the order of service in starting a game may not necessarily be decided by the result of a toss.
There are two basic ways of handling a paddle, vertical style and horizontal style. It may be safe to say that horizontal style players have an advantage in back hand skills while vertical style players have an disadvantage in speed in terms of reaction.

A good player needs a good paddle.

II. Multiple choices: (One correct=1; One incorrect=-0.25; No answer=0).

1. The table shall be in a rectangular surface, 9 ft. in length, and ____ cm. in width.
   - A. 152
   - B. 46
   - C. 152.5
   - D. None of the above.

2. The table shall be made of any material and shall yield a uniform bounce of not less than 22 cm. and not more than ____ cm. when a standard ball is dropped from a height of 30.5 cm. above its surface.
   - A. 24
   - B. 24.25
   - C. 25
   - D. 25.25

3. Components of an attack skill contain: skill form, speed, spin, and ball placement. Which component is essential among the four?
   - A. Skill form
   - B. Speed
   - C. Spin
   - D. Ball placement.

4. The net shall be suspended by a cord attached at each end to an upright post ____ cm. high.
   - A. 15.35
   - B. 15.25
   - C. 15.15
   - D. None of the above.

5. The ball shall be spherical, with a diameter of ____ mm.
   - A. 28
   - B. 38
   - C. 48
   - D. None of the above.
6. Which of the following is not true? The paddle may be of any:

7. A ____ is the period during which the ball is in play.
   A Let   B Play   C Volley   D Rally

8. In a game, the two players (or two sides) may change ends when the score is _____.
   A. 10 to 11   B. 10 to 12   C. 4 to 10   D None of the above.

9. It is possible that a ball served may not possess _____.
   A Speed   B Spin   C Landing point   D None of the above.

10. A skill form of chopping is more like a serve of _____.
    A Topspinning   B Left-to-Right side spinning   C Right-to-Left side spinning   D None of the above
APPENDIX E

DEFINITIONS FOR GAME PLAY RECORDING
Definitions for game play recording.

Forehand Serve: refers to a serve done with the point-finger-side of the paddle for horizontal handling or hit from the right side of the body of a vertical-racket-handling player.

Forehand Topspinning Serve: refers to a forehand serve with topspin.

Forehand Sidespinning Serve: refers to a forehand serve with sidespin.

Forehand Underspinning Serve: refers to a forehand serve with underspin.

Backhand Serve: refers to a serve done with the thump-side of paddle for a horizontal-racket-handling player or hit from the left side of the body of a vertical-racket-handling player.

Backhand Topspinning Serve: refers to a backhand serve with topspin.

Backhand Sidespinning Serve: refers to a backhand serve with sidespin.

Backhand Underspinning Serve: refers to a backhand serve with underspin.

A Good Serve: refers to a legal and good serve which is returned successfully.
A Serve Scored: refers to a legal and good serve which
  gains a point directly.

A Serve Missed: refers to a serve which causes the
  server to lose a point directly.

A rally: refers to the period during which the ball is
  in play.

A Rally Scored: refers to a rally where the target
  player gains a point.

A Rally Lost: refers to a rally where the target player
  loses a point.
APPENDIX F

15-ITEM QUESTIONNAIRE ON ATTITUDE
The following questions are related to the grading system used in this Table Tennis Class. Please make your choice on the right side of each question by circling one of the five: SA=Strongly Agree, A=Agree, N=Neutral, DA=Disagree, SD=Strongly Disagree.

1. I liked this grading system.
2. This grading system was explained clearly.
3. This grading system was fair.
4. Under this grading system, I learned skills better.
5. Under this grading system, I learned more about table tennis.
6. This grading system reflected a true understanding of my capabilities.
7. This grading system differentiated among students.
8. This grading system was consistent with course objectives.
9. Under this grading system, it was easier for me to achieve the grade I wanted.
10. Under this grading system, I worked to achieve my desired grade.
11. This grading system provided guidance for learning.
12. I would prefer that more physical education classes use this kind of grading system.
13. This grading system reflected course content learning expectations. SA A N DA SD

14. This grading system was fair and realistic. SA A N DA SD

15. I recommended this grading system for physical activity courses. SA A N DA SD
APPENDIX G

MEAN COMPARISONS ON ATTITUDE QUESTIONS
Table 47

Mean Comparisons of Student Perceptions on the Three Grading Systems

<table>
<thead>
<tr>
<th>Question</th>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>2.35</td>
<td>1.15</td>
<td>1</td>
<td>4</td>
<td>S-M *</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>1.43</td>
<td>0.85</td>
<td>1</td>
<td>4</td>
<td>S-T *</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>3.33</td>
<td>1.29</td>
<td>2</td>
<td>5</td>
<td>M-T *</td>
</tr>
<tr>
<td>Question</td>
<td>S</td>
<td>2.29</td>
<td>0.61</td>
<td>2</td>
<td>4</td>
<td>S-M *</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>1.21</td>
<td>0.58</td>
<td>1</td>
<td>3</td>
<td>S-T</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>2.53</td>
<td>1.30</td>
<td>1</td>
<td>5</td>
<td>M-T *</td>
</tr>
<tr>
<td>Question</td>
<td>S</td>
<td>2.50</td>
<td>1.16</td>
<td>1</td>
<td>5</td>
<td>S-M *</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>1.29</td>
<td>0.61</td>
<td>1</td>
<td>3</td>
<td>S-T</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>3.46</td>
<td>1.25</td>
<td>1</td>
<td>5</td>
<td>M-T *</td>
</tr>
<tr>
<td>Question</td>
<td>S</td>
<td>2.71</td>
<td>0.99</td>
<td>2</td>
<td>5</td>
<td>S-M *</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>2.00</td>
<td>1.04</td>
<td>1</td>
<td>4</td>
<td>S-T *</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>3.53</td>
<td>1.19</td>
<td>1</td>
<td>5</td>
<td>M-T *</td>
</tr>
</tbody>
</table>
Table 47 (cont')

Question 5. "Under this grading system, I learned more about table tennis."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>2.43</td>
<td>0.76</td>
<td>1</td>
<td>4</td>
<td>S-M</td>
</tr>
<tr>
<td>M</td>
<td>2.00</td>
<td>0.88</td>
<td>1</td>
<td>4</td>
<td>S-T  *</td>
</tr>
<tr>
<td>T</td>
<td>3.47</td>
<td>1.19</td>
<td>2</td>
<td>5</td>
<td>M-T  *</td>
</tr>
</tbody>
</table>

Question 6. "This grading system reflected a true understanding of my capabilities."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>2.93</td>
<td>1.07</td>
<td>2</td>
<td>5</td>
<td>S-M</td>
</tr>
<tr>
<td>M</td>
<td>2.36</td>
<td>1.15</td>
<td>1</td>
<td>5</td>
<td>S-T  *</td>
</tr>
<tr>
<td>T</td>
<td>3.93</td>
<td>1.03</td>
<td>2</td>
<td>5</td>
<td>M-T  *</td>
</tr>
</tbody>
</table>

Question 7. "This grading system differentiated among students."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>2.71</td>
<td>0.73</td>
<td>2</td>
<td>4</td>
<td>S-M</td>
</tr>
<tr>
<td>M</td>
<td>3.29</td>
<td>1.14</td>
<td>2</td>
<td>5</td>
<td>S-T</td>
</tr>
<tr>
<td>T</td>
<td>3.33</td>
<td>1.35</td>
<td>1</td>
<td>5</td>
<td>M-T</td>
</tr>
</tbody>
</table>

Question 8. "This grading system was consistent with course objectives."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>2.43</td>
<td>0.85</td>
<td>1</td>
<td>4</td>
<td>S-M  *</td>
</tr>
<tr>
<td>M</td>
<td>1.79</td>
<td>0.70</td>
<td>1</td>
<td>3</td>
<td>S-T</td>
</tr>
<tr>
<td>T</td>
<td>3.07</td>
<td>1.03</td>
<td>2</td>
<td>5</td>
<td>M-T  *</td>
</tr>
</tbody>
</table>
Table 47 (con't)

Question 9. "Under this grading system, it was easier for me to achieve the grade I want."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>3.14</td>
<td>1.10</td>
<td>2</td>
<td>5</td>
<td>S-M *</td>
</tr>
<tr>
<td>M</td>
<td>1.29</td>
<td>1.07</td>
<td>1</td>
<td>5</td>
<td>S-T</td>
</tr>
<tr>
<td>T</td>
<td>3.80</td>
<td>1.08</td>
<td>1</td>
<td>5</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

Question 10. "Under this grading system, I worked to achieve my desired grade."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>2.79</td>
<td>0.89</td>
<td>2</td>
<td>4</td>
<td>S-M *</td>
</tr>
<tr>
<td>M</td>
<td>1.57</td>
<td>0.94</td>
<td>1</td>
<td>4</td>
<td>S-T</td>
</tr>
<tr>
<td>T</td>
<td>2.93</td>
<td>1.22</td>
<td>1</td>
<td>5</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

Question 11. "This grading system provided guidance for learning."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>2.64</td>
<td>0.84</td>
<td>1</td>
<td>4</td>
<td>S-M *</td>
</tr>
<tr>
<td>M</td>
<td>1.93</td>
<td>0.73</td>
<td>1</td>
<td>3</td>
<td>S-T *</td>
</tr>
<tr>
<td>T</td>
<td>3.47</td>
<td>1.06</td>
<td>2</td>
<td>5</td>
<td>M-T *</td>
</tr>
</tbody>
</table>

Question 12. "I prefer that more physical education classes use this kind of grading system."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>2.86</td>
<td>1.10</td>
<td>1</td>
<td>5</td>
<td>S-M *</td>
</tr>
<tr>
<td>M</td>
<td>1.50</td>
<td>1.16</td>
<td>1</td>
<td>5</td>
<td>S-T *</td>
</tr>
<tr>
<td>T</td>
<td>4.2</td>
<td>0.77</td>
<td>3</td>
<td>5</td>
<td>M-T *</td>
</tr>
</tbody>
</table>
Table 47 (cont)

Question 13. "This grading system reflected course content learning expectations."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
</table>
| S     | 2.43 | 0.85| 1       | 4      | S-M *
| M     | 1.71 | 0.99| 1       | 4      | S-T *
| T     | 3.40 | 1.18| 1       | 5      | M-T *

Question 14. "This grading system was fair and realistic."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
</table>
| S     | 2.57 | 0.94| 1       | 5      | S-M *
| M     | 1.50 | 1.09| 1       | 5      | S-T *
| T     | 3.40 | 1.12| 1       | 5      | M-T *

Question 15. "I recommended this grading system for physical activity courses."

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Test</th>
</tr>
</thead>
</table>
| S     | 2.86 | 1.17| 1       | 5      | S-M *
| M     | 1.50 | 1.16| 1       | 5      | S-T *
| T     | 4.07 | 0.96| 3       | 5      | M-T *

Notes: (a) Sample size for class 1 = 14, for class 2 = 14, and for class 3 = 15.
(b) "Test" = Kruskal-Wallis Tests (Chi-Square Approximation).
(c) "*" = Significance at alpha = .05.